

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 20:06:00 ; Search time 3647 Seconds
(without alignments)
10916.545 Million cell updates/sec

Title: US-09-943-108a-1

Perfect score: 1368

Sequence: 1 atgcattgaaggcttacc.....gtatgaattaccgttttaa 1368

Scoring table: IDENTITY NUC

Gap 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.fun.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1368	100.0	1368	6	E36050	E36050 Signal reco
2	1368	100.0	303750	1	AP003133	AP003133 Staphyloc
3	1368	100.0	347235	1	AP003361	AP003361 Staphyloc
4	1338.4	99.3	295350	1	AP004826	AP004826 Staphyloc
5	790.4	57.8	792	6	E36051	E36051 Signal reco
6	700.8	51.2	319630	6	AX413016	AX413016 Sequence
7	700.8	51.2	349980	6	AX417046	AX417046 Sequence
8	697.6	51.0	2256	6	AX416333	AX416333 Sequence
9	622.8	45.5	2739	1	D14356	D14356 Bacillus su
10	622.8	45.5	208780	1	BSUB0009	299112 Bacillus su
11	567	41.4	11023	1	AE006391	AE006391 Lactococc
12	564	41.2	303249	1	AP001515	AP001515 Bacillus
13	522.8	38.2	296750	1	AP003191	AP003191 Clostridi
14	499.2	36.5	10173	1	AE010040	AE010040 Streptoco
15	497.2	36.3	4152	1	U88582	U88582 Streptococ
16	496	36.3	1569	6	AR081886	AR081886 Sequence
17	496	36.3	1569	6	AR195303	AR195303 Sequence
18	496	36.3	1569	6	E35383	E35383 Novel ffh.
19	496	36.3	1605	6	AX194269	AX194269 Sequence
20	496	36.3	7577	6	BD003726	BD003726 Polynucle
21	496	36.3	9985	1	AE007428	AE007428 Streptoco
22	496	36.3	10462	1	AE008489	AE008489 Streptoco
23	496	36.3	50946	1	AE014153	AE014153 Streptoco
24	492.8	36.0	1572	6	AX194043	AX194043 Sequence
25	492.8	36.0	77743	2	SPNEU1910	AL449932 Streptoco
26	491.2	35.9	12187	1	AE006560	AE006560 Streptoco
27	486.4	35.6	1136	6	AX433432	AX433432 Sequence
28	479	35.0	11024	1	AE007684	AE007684 Clostridi
29	465.4	34.0	11825	1	AE010644	AE010644 Fusobacte
30	443.4	32.4	10574	1	U32696	U32696 Haemophilus
31	435.4	31.8	11179	1	AE014114	AE014114 Buchnera
32	434.4	31.8	14571	1	AE013104	AE013104 Thermosma
33	433.8	31.7	11632	1	AE006158	AE006158 Pasteurel
34	424.2	31.0	11004	1	AE013692	AE013692 Versinia
35	424.2	31.0	220050	1	A7414156	A7414156 Versinia
36	418.2	30.6	293181	1	AP001119	AP001119 Buchnera
37	416.6	30.5	4371	1	BAP239043	AJ239043 Buchnera
38	409.2	29.9	10566	1	AE004142	AE004142 Vibrio ch
39	404	29.5	92407	6	AX067461	AX067461 Sequence
40	404	29.5	343550	1	AP003587	AP003587 Nostoc sp
41	394.2	28.8	2817	1	MYCSRPW54A	X91593 Mycoplasma
42	393.6	28.0	321250	1	MPULM02	U1445564 Mycoplasma
43	377.4	27.6	143308	1	D90913	D90913 Synecocyst
44	374.4	27.4	11081	1	AE002108	AE002108 Ureaplasma
45	370.6	27.1	4586	1	ECTRMD	X01818 E. coli trm

ALIGNMENTS

RESULT 1
E36050
LOCUS E36050 1368 bp DNA linear PAT 18-JUN-2001
DEFINITION Signal recognition particle polypeptide and polynucleotide.
ACCESSION E36050
VERSION E36050.1 GI:13022452
KEYWORDS JP 1999235183-A/1.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Michael,T.B.
TITLE Signal recognition particle polypeptide and polynucleotide
JOURNAL Patent: JP 1999235183-A 1 31-AUG-1999;
SMITHKLINE BEECHAM CORP

COMMENT	OS	Unidentified
	PN	JP 1999235183-A/1
	PD	31-AUG-1999
	PF	03-SEP-1998 JP 1998289963
	PR	03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 P1
	MICHAEL TERENCE BLACK	
	PC	C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00,
	PC	A61K31/00,
	PC	A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,
	PC	A61K39/085,
	PC	A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,
	PC	C12N1/19,
	PC	C12N1/21, C12N5/10, C12N5/10, C12P21/02, C12Q1/68, G01N33/15, PC
	G01N33/566//	
	PC	C12P21/08, G01N33/577, (C12N15/09, C12R1:445), C12N5/00, C12N5/00,
	PC	C12N5/00,
	PC	(C12N15/00, C12R1:445)
	CC	Strandedness: Single;
	CC	Topology: Linear;
	FE	Key
	FT	source
	FT	Location/Qualifiers
FEATURES	Location/Qualifiers	
source	1..1368	
	/organism="Unidentified".	
	/db_xref="taxon:32644"	
BASE COUNT	521 a	186 c 288 g 373 t
ORIGIN		
Query Match	100.0%; Score 1368; DB 6; Length 1368;	
Best local similarity	100.0%; Pred. No. 5.3e-220;	
Matches 1368; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGGCATTTGAAGGCTTATCAGACGCTCGCAGCGAGATGCAAAATTCGGTGAAG 60
DB	1	ATGGCATTTGAAGGCTTATCAGACGCTCGCAGCGAGATGCAAAATTCGGTGAAG 60
QY	61	GGTAACTTACTGAAGCTGATTAAGATAATGATGCGTGAAGTAAGATGAGCTTATT 120
DB	61	GGTAACTTACTGAAGCTGATTAAGATAATGATGCGTGAAGTAAGATGAGCTTATT 120
QY	121	GAGGCTGAGGTAACTTTAAAGTGGTAAAGATTTATTAACACAGATCAGAACCGCA 180
DB	121	GAGGCTGAGGTAACTTTAAAGTGGTAAAGATTTATTAACACAGATCAGAACCGCA 180
QY	181	TTAGGTTCCGATGTAATCAATCATTAACACAGGCAAGTATTAAATAGTTCAA 240
DB	181	TTAGGTTCCGATGTAATCAATCATTAACACAGGCAAGTATTAAATAGTTCAA 240
QY	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATPACATCGATTAATATGTCAAATAACCA 300
DB	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATPACATCGATTAATATGTCAAATAACCA 300
QY	301	CCACTGTTGTTATGATGGTGGTTTACAGTGGCTGGTAAACACACAGCTGAGTAA 360
DB	301	CCACTGTTGTTATGATGGTGGTTTACAGTGGCTGGTAAACACACAGCTGAGTAA 360
QY	361	TTAGCATTTATGATCGTAAAAATACACAAAAACCTATGTTAGTTGAGCAGATATT 420
DB	361	TTAGCATTTATGATCGTAAAAATACACAAAAACCTATGTTAGTTGAGCAGATATT 420
QY	421	TATGCTCCAGCGGATTAATCAATTAACACAGTAGGGAACAAATGTATATTCCTGTA 480
DB	421	TATGCTCCAGCGGATTAATCAATTAACACAGTAGGGAACAAATGTATATTCCTGTA 480
QY	481	TACAGTGAAGGATCAAGTAAAGCCACACAAATTTGAATCAATTAACATGCT 540
DB	481	TACAGTGAAGGATCAAGTAAAGCCACACAAATTTGAATCAATTAACATGCT 540
QY	541	AAAGAGACATTTAGCTTTGATCATGATACAGAGGTCGATACACATCGATGAA 600
DB	541	AAAGAGACATTTAGCTTTGATCATGATACAGAGGTCGATACACATCGATGAA 600

QY	601	GCATTGATGAACGAATTAAGAAAGTAAAGAAATTCCTAAACCAACGAAATATGTTA 660
DB	601	GCATTGATGAACGAATTAAGAAAGTAAAGAAATTCCTAAACCAACGAAATATGTTA 660
QY	661	GTTGTCGATCAATGAACGGGTCAAGATGCTGTCAATGTTCAGAAATCTTTTACGATCAA 720
DB	661	GTTGTCGATCAATGAACGGGTCAAGATGCTGTCAATGTTCAGAAATCTTTTACGATCAA 720
QY	721	CTTGATGTCACAGGTGTTCTTAACTTAATTAAGATAGTGTGATACACGTGGTGGTCACT 780
DB	721	CTTGATGTCACAGGTGTTCTTAACTTAATTAAGATAGTGTGATACACGTGGTGGTCACT 780
QY	781	TTATCTATTTCGTTGGTGCACACAAAAACCAATTAATTTGTTGTTAGTGAAGAAAGTTA 840
DB	781	TTATCTATTTCGTTGGTGCACACAAAAACCAATTAATTTGTTGTTAGTGAAGAAAGTTA 840
QY	841	GATGTTTATGAGCTATTTCATCTGACGATATGGCATCAGCATTTTAGTATGGGTAT 900
DB	841	GATGTTTATGAGCTATTTCATCTGACGATATGGCATCAGCATTTTAGTATGGGTAT 900
QY	901	GTCTTAAGTTTAAATTTGAAAAGCGCACACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960
DB	901	GTCTTAAGTTTAAATTTGAAAAGCGCACACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960
QY	961	GAGAAAAAGATGCGTGAAGTCAATCGTTTACTTTAGATGATTTTITAGAACAACTTGATCAG 1020
DB	961	GAGAAAAAGATGCGTGAAGTCAATCGTTTACTTTAGATGATTTTITAGAACAACTTGATCAG 1020
QY	1021	GTCAAAATCTAGGACCACTGGATGATATTTATGAAAATGATCCAGGTATGAATAAATG 1080
DB	1021	GTCAAAATCTAGGACCACTGGATGATATTTATGAAAATGATCCAGGTATGAATAAATG 1080
QY	1081	AAAGGCTAGTAAGCTTAAATGATGAGTGAAGAAAGCAAAATGATCATATAAAGCGATTATC 1140
DB	1081	AAAGGCTAGTAAGCTTAAATGATGAGTGAAGAAAGCAAAATGATCATATAAAGCGATTATC 1140
QY	1141	CATCATGACCGCGCTGAAAGAACATCCAGACACATTTGATGATGATCACTAATGATCACTAATAAAG 1200
DB	1141	CATCATGACCGCGCTGAAAGAACATCCAGACACATTTGATGATGATCACTAATGATCACTAATAAAG 1200
QY	1201	CGTATTGCTAAAGGCTGCTGCTTCATTACAGAGTCAATCGTTTGTGATGAACAAATTT 1260
DB	1201	CGTATTGCTAAAGGCTGCTGCTTCATTACAGAGTCAATCGTTTGTGATGAACAAATTT 1260
QY	1261	AACGATATGAAGAAATGATCAACAAATTCACCTGGTGGCGTAAAGGTAAAAAGGTAAA 1320
DB	1261	AACGATATGAAGAAATGATCAACAAATTCACCTGGTGGCGTAAAGGTAAAAAGGTAAA 1320
QY	1321	CGCAATCAATGCCAAATATGTTTAAAGGTATGAATTTACCGTTTAA 1368
DB	1321	CGCAATCAATGCCAAATATGTTTAAAGGTATGAATTTACCGTTTAA 1368
RESULT 2		
AP003133	LOCUS	303750 bp DNA linear BCT 02-JUL-2002
Staphylococcus aureus subsp. aureus N315 genomic DNA, complete genome, section 5/10.		
AP003133	BA000018	ACCESSION
AP003133.2	GI:14349175	VERSION
Bacteria; Firmicutes; Bacillales; Staphylococcus.		
Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315, strain:N315) DNA.		
Staphylococcus aureus subsp. aureus N315		
Bacteria; Firmicutes; Bacillales; Staphylococcus.		
1		
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Di, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.		

TITLE Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
PUBMED 11418146
REFERENCE 2 (bases 1 to 303750)
AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K. and Kikuchi, H.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center, 2Chome 49-10 Nishihara, Saitama-ku, Tokyo 151-0066, Japan
COMMENT (E-mail: oguchi@nitech.go.jp, URL: <http://www.bio.nite.go.jp/>,
 Tel: 81-3-3481-8423, Fax: 81-3-3481-8424)
FEATURES On Jun 12, 2001 this sequence version replaced gi:13701012.
 Location/Qualifiers
 1. 303750
 /organism="Staphylococcus aureus subsp. aureus N315"
 /strain="N315"
 /sub_species="aureus N315"
 /db_xref="taxon:158879"
 202. .1155
 /gene="SA1056"
 202. .1155
 /gene="SA1056"
 /feature="ORFID:SA1056"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAB42308.1"
 /db_xref="GI:13701013"
 /translations="MKKTGCLLIMLVVAGSGFGNGHKLSSKSESKEQETVKRES
 EEEDPLEXYEIEKMGKIGKADPSLDKLDPLMTEKSTNSKGIQGEKQYKELMGKY
 ELADYRTKRGSSIKQVDAFFAKGKGIARKVIEIHDHVDVAQVDIWIYDDEPKKIGNS
 NTPFYAEIMTKYDKGLVYASVEPGSVIHKDDAIKYDDYSKLKLSQLTKLDHKKP
 IYVDAVQIKFGVPLTSVFMTHGSKDTKDEVLPALEYTFSPKNYEDKSNPDPKVLN
 LVHDFNASSDFGNARFVLSKVIKESYNETASDLSLK"
 /gene="SA1057"
 complement(1318. .1521)
 /gene="SA1057"
 /feature="ORFID:SA1057"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAB42309.1"
 /db_xref="GI:13701014"
 /translations="MQYNKVIYILLALFLSGFGIHKFYAKKNMQGILHLIFCWGIPH
 ILALISAVITIVFKPDAEQGNVTL"
 1820. .2308
 /gene="SA1058"
 1820. .2308
 /gene="SA1058"
 /feature="ORFID:SA1058"
 /hypothetical protein, similar to polypeptide deformylase"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB42310.1"
 /db_xref="GI:13701015"
 /translations="MAIKKLVPAHPILTKKAQAVTKFDLSLKLQLEDITWYAOEA
 AGLCAPQNSLOVALIDMEGGLQVNPKITLSQNETITDLEGSTLPLPYKGEVTR
 SKAVVVSVDVNGKVELPAHEDVARMILHILIQMGNGIPTTERADRLITDKEVAYFI
 ND"
 2301. .3236
 /gene="SA1059"
 2301. .3236
 /gene="SA1059"
 /feature="ORFID:SA1059"
 /codon_start=1
 /transl_table=11
 /product="methionyl-tRNA formyltransferase"
 /protein_id="BAB42311.1"
 /db_xref="GI:13701016"

/translation="MTKILPMGPDPFSTTVLEMLIAEHDVIAVVTQDPFVGRKRVMT
 PPVKKVAKYDLFVYQPEKLSGELEQLQLOLDVLIYTAARFQQLLPESLIAPEKLQ
 AINVHSLPLRYGGAPEIHOAIIDGEOETGITIMVKKLDAGNIISQOAIKEEND
 VGTMDKLSVLGADLLKETLPSIIEGTNESVPQDQTQATFASIRREDERISWKPGR
 OVFNQIRGLSPWPVAYTTMDTNLKYDAELVETNKINEPFTIETTKKAIIVATND
 BAYAKDMOLAGKRMALAAVLSGAQNTLVGKKLI"
 3155. .4540
 /gene="SA1060"
 3155. .4540
 /gene="SA1060"
 /feature="ORFID:SA1060"
 /hypothetical protein, similar to RNA-binding Sun protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB42312.1"
 /db_xref="GI:13701017"
 /translation="MSWEKENVSCQLFKWCAKHTSREETYMINVRSIAFDITQDILN
 EGAYSNLRINEVLSNELNADKALETVIGTVKRYTLDVFLKFPVTKTKAWVRQ
 LLWMSYQVYLDKVPNEAIIINEAVEIAKERGGYHGVNGVNGILRTMBSDLDPDNEI
 ADPKMALEYSMFKWIIDHWAHYGLETEETILQSTLETSTVTVANLTRASLDII
 EKLQGEYDVERKDHLDPLCHIGGQPIIHSRSFDDGFVSIDKSSMFVAHIMNVDHND
 HVLDACSPGGKACHIAEVLMPGQYDASDIDHDKIDILINFTIKKLRITNKAFOHDA
 TRPYDKYDKILVDAPCGSLGVNMRHKPEIKYTSQKHIESLVELQLEILENVKKNYKI
 GGEIYVSTCTEQLNENVIYTELKNNKNEFEPEFQHPITGELVKTQIMPQDENSOG
 FRTIKTKRKN"
 4543. .5637
 /gene="SA1061"
 4543. .5637
 /gene="SA1061"
 /feature="ORFID:SA1061"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAB42313.1"
 /db_xref="GI:13701018"
 /translation="MITAEKKKKKFLPNFDKQSIYSIRFDEMQLVLEQGGQKFRPK
 QIPFELYQKRVDSIDENLNSKDLROLLDNFTVTTLTVVQESKDGITKELFELQD
 GYTIETVLRMDHGVNSVTVQVCRIGCTFCAGTLGKRLNEAGEIYVSQVLTFOKA
 LDATERSYQIVIMGIGEPFENDEMDFLRIVDNDNSLNGARHTVSTGIIPIRY
 DRADEDIQINFAVSHAAKDEVRSLMPINRAVAVKELIAIQYQYQKINRRTVEYG
 LFGVNDQLHARELAHLIKLNCNVNLIPIVNHVPERNYVTKANDIFRFEKELARLG
 INATIRREGSGSDAACQLRAKROVETR"
 5644. .6387
 /gene="SA1062"
 5644. .6387
 /gene="SA1062"
 /feature="ORFID:SA1062"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAB42314.1"
 /db_xref="GI:13701019"
 /translation="MLEAQPTDTGHRKDNEDAGGIFYNOTNQLILVLDGMMGKHA
 GEVASFVDEKLSFEAEENLIEHOAEWLNRLNKLIDNLFOLYHGAENYKGMGT
 CVCALVFESVLIANGDSRAYVINSKQLEQITSDHSVAHLVLTQGITPEAPFPEZO
 RNIITKMGTDKRVSPDLTKRLNFIDYLLNSDGLTDYVDKNEIKRLLVKEGTIEDH
 GDQLMQLADNHSKDNVTFILAALEGDKV"
 6384. .8378
 /gene="SA1063"
 6384. .8378
 /gene="SA1063"
 /feature="ORFID:SA1063"
 /codon_start=1
 /transl_table=11
 /product="protein kinase"
 /protein_id="BAB42315.1"
 /db_xref="GI:13701020"
 /translation="MIGKILNERYKIVDKLGGGMSTVYLAEDTILNKAIVKALFIP
 PREKETLARPEREYHNSQLSHQNIIVMIDVDEDDCYILWMEYIEGTPLSEYESH
 GPLSVDTAINTNQILDKHAHMKRIVHRDIKPNILIDNSNLTKLIFDFGLAKALSE
 TSLTQTNHVLGTVQYFSEPAQKEATDECTDIYSIGVLYEMLYGPPFNGPAGVYSIA
 IKHIQDSVPNVTDRKDIQPSLVILRATEKDKANRYKTIQEMKDDLSVSLHENRA
 NEDVYELDKMKTIAVPLKEDLAKHISEHKSNOQPKRETTQVPVINGFAHQOQFQKEG

REFERENCE
AUTHORS

1 Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I., Cui L., Oguchi A., Aoki K., Nagai Y., Lian J., Ito T., Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani O., Takahashi N., Sawano T., Inoue R., Kaito C., Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M., Ogasawara N., Hayashi H. and Hiramatsu K. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*

TITLE

Lancet 357 (9264), 1225-1240 (2001)

JOURNAL

MEDLINE

21311952

REFERENCE

2 (bases 1 to 347235)

AUTHORS

TITLE

Direct Submission
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College of Medical Technology and Nursing, Department of Medical Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail: tohtesakura.cc.tsukuba.ac.jp, Tel: 81-298-53-3434, Fax: 81-298-53-3454)

COMMENT

On May 29, 2001 this sequence version replaced gi:13875305.

FEATURES

SOURCE

1. .347235
/organism="Staphylococcus aureus subsp. aureus Mu50"

/strain="Mu50"

/sub_species="aureus Mu50"

/db_xref="taxon:158878"

35. .1021

/gene="SAV0992"

/gene="SAV0992"

/codon_start=1

/transl_table=11

/product="oligopeptide ABC transporter ATP-binding protein homologue"

/protein_id="BAB57154.1"

/db_xref="GI:14246762"

/translation="MNNVLEVKDLETSLKINNEWLATVENISPELSKGEVIGVSGCGSKLSIKLPEKISLSSGEVIFDGRKIDTFLNEKQLDINGNDAMIFQPPM TALNPETIKQLVESIKSHKIKSKKANLAKDLKVGTAQDEILNSYPOLSGM MRQWPTAMAISCPKLAFADEPTALDVTIQAILDLKELQETQWAMIMTHDLS VVAFCDKLVVMYAGQIVFEGGKEILHNEKHPYTKLILSTPKLKEGQKRLFTIBGI VPSTQAPHVKNCFRANCKNKLDICNQSFKMHCVEDVIVRCHLVKNEYKEI"

1024. .2004

/gene="apbf"

1024. .2004

/gene="apbf"

/note="SAV0993"

/codon_start=1

/transl_table=11

/product="oligopeptide transport system ATP-binding protein Apbf homologue"

/protein_id="BAB57155.1"

/db_xref="GI:14246763"

/translation="MENILEVNQIKVYKIKTGLOKTOYKAVDVSFSIKKGOTFG LVGSGGKSTGLKGVITRLEADGTVIRLGKIKLRKSRQOQIMFDQYVA SLNPMQWGDIISEPLIYKIKPEIKKVLKCYGLSEDAVYKXAFESGGRQ RVGIARALARPLIVADPVSALDVSQVNLNKLKDLQGFKLSTYFAHDLISVK HISVIGWLVGHIVETASDKEIYENPKHPTKALISSIPOIDKNNRRIILGELPS PSNPQCPPTTCPIAKDKCKENIPOLKDIGNEHQVACFYVDYKVDLNA"

1997. .2959

/gene="oppB"

1997. .2959

/gene="oppB"

/note="SAV0994"

/codon_start=1

/transl_table=11

/product="probable oligopeptide transport system permease protein"

/protein_id="BAB57156.1"

/db_xref="GI:14246764"

/translation="MLKLIKRLGLMIFLLILISIVFSLAIQDFPDLQNGIK QBAINAQREKLGNDSTPHQYIRWNVHVIHGDLSGESIKYKRPVIDVIERIPNTILIG

gene

CDS

gene

CDS

AMSLIITYISFALGITSGRYSLSYTDYTVQIFNYLMLAIPSTIAGVFAIFISPELQ WPFQGSVDNLKGETFEYTKSYHTFLPALTLGLSTAGYIOYLRNDIENSKDY VLFARSKLNMKNKYNKHLRNSLIPLITFLGADIVSLTGAVTIFPSINGICKLF LESVIGQDYLPMALITLFFSFLGLNLISDIITYGFIDPRISN"

2971. .3852

/gene="SAV0995"

2971. .3852

/gene="SAV0995"

/note="similar to oligopeptide transport system permease protein OppC"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57157.1"

/db_xref="GI:14246765"

/translation="WQNKSKPFIKALSRFHINKTAMLSIIFALITIVSIITAPIAP FPNODLNLKGETFAONILGDSGRDNFSELIYAGRISSISIGTSIGMLIGIT VGVISYFGIVDITLMRIEFTVLPFLIFALIVNAALGDKINPYGSAIILVLII VLSWGIALRVNGKVLQEKENEYFLAAKSIGTPTTKIILKHLNLSLVVIVQATLFF AGMIVVSGLSFLGFGISKAIPSGNMLSDAQEGDVISGKFWIWPPIAMITLISL NFVGEGLKDAFNPRGR"

complement(3894. .4883)

/gene="trpS"

complement(3894. .4883)

/gene="trpS"

/note="SAV0996"

/codon_start=1

/transl_table=11

/product="tryptophanyl-tRNA synthetase"

/protein_id="BAB57158.1"

/db_xref="GI:14246766"

/translation="METHFGISQPIPTIGNVIGALKQFVDVQNDYDCEVCIVDQHA ITMPDRLKLRQTFQAAIYLAGSDIDPKATLFIQSEVPAHVQAGMLTITASVGEEL RNSQYDQKRAQVAGIYLPFLMAADIVLNTNIVPVGDDQKHTELTNRNLD RFNSRYNDVLMPEIRMEKVGGRVMSLQDPTKMSKSDNAKNTISLLDEPNVAAKI KSAVTDSDGIIKFRDNKPGTTLISYAGLTDMPDKDIEAKYEGEGYKFGKDLAEI VKAFIVFEQEKYSEFYNDSKDLDDILQDGRKAHKVSEKTVKKMEKAGLGRKR"

5177. .5572

/gene="SAV0997"

5177. .5572

/gene="SAV0997"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB57159.1"

/db_xref="GI:14246767"

/translation="MVTLFPSCTSCRKAKAWLQEHDPYTERNIFSEHLTIDEIKQ ILKMTEDGTEIISRSKTYQKLNVDISLPLQDLYSIQDNPGILRRPITLDNKRLO VGYNEIDIRFLPRKRVTFQLEAQRMD"

5943. .6662

/gene="SAV0998"

5943. .6662

/gene="SAV0998"

/note="similar to negative regulator of genetic competence MeoA"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB57160.1"

/db_xref="GI:14246768"

/translation="NRIERVDDTVTKLFITYSDIARGFSREDLWTKRKEGFWSM MDEINDEEDFVVEGLPWIOVHAPKGVETISKKNEDMMMSDDDATDQFQOVEL LAOTLEGDQLEPEQTEKKEKQSKRKQSKARNTRTITIVAFNDELVINYAVHS NPITTEFEDLLKWDGTYTYAYAFHDSHVQDEVINDYSQLLERAYPIDRTEVINDYA KIIMSHNVIAQVRRYPETTE"

6783. .7769

/gene="SAV0999"

6783. .7769

/gene="SAV0999"

/note="similar to transcription factor"

/codon_start=1

/transl_table=11

/product="hypothetical protein"


```

strain:MW2, section 5/10.
ACCESSION AP004826 BAO000033
VERSION AP004826.1 GI:21204263
KEYWORDS
SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.
ORGANISM Staphylococcus aureus subsp. aureus MW2
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Baba.T., Takeuchi.F., Kuroda.M., Yuzawa.H., Aoki.K., Oguchi.A.,
Nagai.Y., Iwano.N., Asano.K., Naimi.T., Kuroda.H., Cui.L.,
Yamamoto.K. and Hiramatsu.K.
Genome and virulence determinants of high virulence
community-acquired MRSA
Lancet 359 (9320), 1819-1827 (2002)
22040717
PUBMED
REFERENCE
2 (bases 1 to 295350)
Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwano,N., Baba,T.,
Kuroda,M., Hiramatsu,K. and Kikuchi,H.
Direct Submission
Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
FEATURES
Source
1. .295350
/organism="Staphylococcus aureus subsp. aureus MW2"
/strain="MW2"
/db_xref="taxon:196620"
301..1254
/gene="MW1096"
301..1254
/gene="MW1096"
/notes="ORFID:MW1096"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94961.1"
/db_xref="GI:21204263"
/translation="MKKTGCLLLMLLVAGCSFGGNHLSKSKSESKQPTVKES
EEKDPLEKEIEKKMGKIDKSLDLPATERSFTNSKIQGWDKYELMGKV
ELADYFTKDSKSSIKDVDAFTKGKKIKRKVIETHDDVDYVWDGDKIGNS
NTVFAETMKYKDKLVZASVPGSVYHKDAAKYDDYSKLKLSOLTLDHPK
VPSVAQIKSCGVELTSVSPMTGSKDTEVLPALEYFESPKNYEDKSNPDPLVN
LVAMDPLASSDFGNHFFVLSKFIKEISNTEHATSDSLK"
/gene="MW1097"
complement(1417..1695)
/notes="MW1097"
/notes="ORFID:MW1097"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94962.1"
/db_xref="GI:21204265"
/translation="MFELKVGIIKREANVIISNSTGKGMQVKNKVIYILLALFSGF
IHKFYAGKNMGILLHFCWTGTPHILAIISAVITVFPKPEQGNVTL"
1919..2407
/gene="MW1098"
1919..2407
/gene="MW1098"
/notes="ORFID:MW1098"
hypothetical protein, similar to polypeptide deformylase"
/codon_start=1
/transl_table=11
/protein_id="BAB94963.1"
/db_xref="GI:21204266"
/translation="NAIKKLPASHPLTKQAQVIKFDLSKLRLQDLEDNVAQEA
AGLCAPQINQLOVALIDMEMEGLQIVNPKIISOSNETITDLEGSLTLPDYGVETR
SKMIVVESDVNGNKVELTAHEDVAMLIHIDQNGIPFTRADRILLTDEKEVYFI
ND"
2400..3335
/gene
2400..3335
/gene="MW1099"
/notes="ORFID:MW1099"
/codon_start=1
/transl_table=11
/product="methionyl-tRNA formyltransferase"
/protein_id="BAB94964.1"
/db_xref="GI:21204267"
/translation="MTKTIEMGTDEFTVLEMLIAEHVDVIAVTPQDRPVGRKVMY
PPPKVKAMKDYLPYQPEKISGSEEEQLLDVLIWTRAFGQLPESLLALPKLG
AINVHASLPKRYGGAPIHQALIDEGQETITIMYVKKLDAGNIISQQAILEEND
VGTMDKLSVLGADLKLETPSIIEGNESVPQDQTAFTASNIRREDINMKNPGR
QVFNQIRGLSPFWAYVTTMDNTNKIYDAELVETKNBEFTIETTKKAIIVTNDN
BAVAIKDMQLAGKKMLAANYLSGAQNTLVGKKLI"
3332..4639
/gene="MW1100"
3332..4639
/gene="MW1100"
/notes="ORFID:MW1100"
hypothetical protein, similar to RNA-binding Sun protein"
/codon_start=1
/transl_table=11
/protein_id="BAB94965.1"
/db_xref="GI:21204268"
/translation="MIENVRSIAFTTQDILNEGAYSNLRINEVLSENEELNMDKALF
TEIYGVTKRKTFLDFYLPKPFYTKIKAWRQLLWMSIYQYVYLDKVPNHAIINEAVF
IAKRGYHNGYVNGILTMRSDLDFNEIADPKRMAIEYSMPKWIIDWATHYG
LEKTEITLQSFLETTSTTVRANLRASLDLDDIIEKLODEGYDVEKDHDLPCLHIGGP
IHSRSFQDFGFSIQDKSSMFVAHIMVYDRHDVLDACAPGKGKACHAEVLMPEGOV
DASDIDHKIDLLINENIKRLTNLKAFOHDAKPYDKTYDKILDVADAPCSLGVMRHK
PEIKYTQSKOHIESLIVLEOLEILENVNKNVKGIELIYSTCTIEQLENEVITFLKN
KNFEFFQHPHITSELVKTQLQIMPQDFNSDGFITKIRKKN"
4642..5736
/gene="MW1101"
4642..5736
/gene="MW1101"
/notes="ORFID:MW1101"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94966.1"
/db_xref="GI:21204269"
/translation="MITAEKKKKKFLPNPKOSIYSLRDEMQWLVQSQOKERAK
QIFENLQKAVDSIDENLNSKDLQLLKNFTVTLTVVKQSKDSTIEFLYELQD
GTTIEVLRHGDIGNSVYTIQVGRIGICIFCAISLGLKRNLEAGEIVSOVLIVQKA
LDATFERSQIVIMGCEPFENYDEMDDFLRVNDNLSNIGARHITVSTSGIIPRIY
LFADEDIOINFAVSLHAAXDEVSRSLMPINRAYNVEKLIEATQYTOEKTRRVTFEY
LFGGVNDQLEHARELAHLIKGLNCHVNLIPYVHPERNYVTKAKNDIPEKFKELRLG
INATIRREGSDIDAACGQLRAKERQVETR"
5743..6486
/gene="MW1102"
5743..6486
/gene="MW1102"
/notes="ORFID:MW1102"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94967.1"
/db_xref="GI:21204270"
/translation="MLEAQFTDTGQHDKNEDAGGIYFNRTNQQLVLCDSNGGHKA
GEVAKFVTDELKSAENLIEHQAENLRNNIKDINFOLYHYAQENAEYKGMGT
CVCAIVFEKSVIANFSDRSRAVYINSRQIQITSDHSFVNHLVLTGQITPEEATPQ
RNIITKMGDTDKVSPDLFIKRLNFDYLLNSDGLTDYVKDNEIKRLVKEGTIEDH
GQQLMQALDNHSDKNVTFILAAIEGKY"
6483..8477
/gene="MW1103"
6483..8477
/gene="MW1103"
/notes="ORFID:MW1103"
hypothetical protein, similar to protein kinase"
/codon_start=1
/transl_table=11

```



```

Db 27724 CGCAATCAATGCAAAATATGTTAAAGAGTATGAATTTACCGTTTAA 27771
|||||
E36051 792 bp DNA linear PAT 18-JUN-2001
LOCUS
DEFINITION Signal recognition particle polypeptide and polynucleotide.
ACCESSION E36051
VERSION E36051.1 GI:13022453
KEYWORDS JP 199235183-A/2.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 792)
AUTHORS Michael,T.B.
TITLE Signal recognition particle polypeptide and polynucleotide
JOURNAL Patent: JP 199235183-A 2 31-AUG-1999;
SMITHKLINE BEECHAM CORP
COMMENT OS Unidentified
PN JP 199235183-A/2
PD 31-AUG-1999
PF 03-SEP-1998 JP 1998289963
PR 03-SEP-1997 US 60/057890,05-MAR-1998 US 09/035382 PI
MICHAE L TERENCE BLACK
PC C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
A61K39/085,A61K39/395,A61K39/395,A61K48/00,C07K14/31,C07K16/12,C12N1/15,
C12N1/19,
PC C12N1/21,C12N5/10,C12N5/10,C12P21/02,C12Q1/68,G01N33/15, PC
G01N33/566//
PC C12P21/08,G01N33/577,(C12N15/09,C12R1:445),C12N15/00,C12N5/00,
C12N5/00,
PC (C12N15/00,C12R1:445)
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..792
FT source /organism='Unidentified'.
FEATURES
source
1..792
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 296 a 116 c 164 g 216 t
ORIGIN
Query Match 57.8%; Score 790.4; DB 6; Length 792;
Best Local Similarity 99.9%; Pred. No. 4.2e-123;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCTTTGAAGCTTATCAGACGCTGCGAGCGAGATGCAAAATATGCGTGAAG 60
Db 1 ATGGCAATTTGAAGCTTATCAGACGCTGCGAGCGAGATGCAAAATATGCGTGAAG 60
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTTATT 120
Db 61 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTTATT 120
QY 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTAATAACAGTATCAGAACGCGCA 180
Db 121 GAGCTGACGTAACATTTAAAGTGGTAAAGAAATTTAATAACAGTATCAGAACGCGCA 180
QY 181 TTAGTTCGGATGTAATCAATCATTAACACAGGCGCAAGTTAATAAATAGTTCAA 240
Db 181 TTAGTTCGGATGTAATCAATCATTTACACAGGCGCAAGTTAATAAATAGTTCAA 240
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATACATCGATTAATATGTCMAATAAACCA 300
Db 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATACATCGATTAATATGTCMAATAAACCA 300
QY 301 CCTACTGTTGTTATGATGGTTTACANGGTGCTGGTAAACAAACAACTGCAGTAA 360

```

```

Db 301 CCTACTGTTGTTATGATGGTTTACAAAGTGGCTGGTAAACAAACAACTGCAGTAA 360
QY 361 TTAGCATTATTGATGCTAAATAATACAAACAAAACCTATCTTATGTTGCAGCAGATAT 420
Db 361 TTAGCATTATTGATGCTAAATAATACAAACAAAACCTATCTTATGTTGCAGCAGATAT 420
QY 421 TATGCTCCACAGCGTAAATCAATTAACAAACAGTAGGGAACAAATATGATATTCCTGTA 480
Db 421 TATGCTCCACAGCGTAAATCAATTAACAAACAGTAGGGAACAAATATGATATTCCTGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACATTAATGCAATTAATAACATGCT 540
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACATTAATGCAATTAATAACATGCT 540
QY 541 AAAGAAGAAACATTAGACCTTGTATCAATGATACAGCAGGTCGATTAACATCGATGAA 600
Db 541 AAAGAAGAAACATTAGACCTTGTATCAATGATACAGCAGGTCGATTAACATCGATGAA 600
QY 601 GCATTGATGACGATTAAGAAAGAAATTTGCTAAACAAATTTGCTAAACAAATTTAGTTA 660
Db 601 GCATTGATGACGATTAAGAAAGAAATTTGCTAAACAAATTTGCTAAACAAATTTAGTTA 660
QY 661 GTTGTCGATTCATGACGGGTCAAGATGCTGCTCAATGCTGCAATCTTTTGACGATCAA 720
Db 661 GTTGTCGATTCATGACGGGTCAAGATGCTGCTCAATGCTGCAATCTTTTGACGATCAA 720
QY 721 CTTGATGTCACAGGTTTACCTTAACATTAAGTAGGTTGATACAGTGGTGGTGCAGCT 780
Db 721 CTTGATGTCACAGGTTTACCTTAACATTAAGTAGGTTGATACAGTGGTGGTGCAGCT 780
QY 781 TTATCTATTCGT 792
Db 781 TTATCTATTCGT 792
RESULT 6
LOCUS AX413016 319630 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser,P. and Kunst,F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 7 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Location/Qualifiers
source
1..319630
/organism='Listeria innocua'
/db_xref='caxon:1642'
BASE COUNT 105207 a 55428 c 66726 g 92263 t
ORIGIN
Query Match 51.2%; Score 700.8; DB 6; Length 319630;
Best Local Similarity 70.7%; Pred. No. 1e-108;
Matches 933; Conservative 0; Mismatches 387; Indels 0; Gaps 0;
QY 1 ATGGCAATTTGAAGGCTTATCAGAACGCTGCAAGCAGCAGATGCAAAATATCGGTGAAG 60
Db 187924 ATGGCAATTTGAAGGCTTATCAGAACGCTGCAAGCAGCAGATGCAAAATATCGGTGAAG 187983
QY 61 GGTAACTTACTGAAGCTGATATAAGTAATGATGCGTGAAGTATGATGCGTGAAGTATGCTTATTT 120
Db 187984 GGTAACTTACTGAAGCTGATATAAGTAATGATGCGTGAAGTATGATGCGTGAAGTATGCTTATTT 120
QY 121 GAGGCTGACGTAACATTTAAAGTGGTAAAGAAATTTAATAACAGTATCAGAACGCGCA 180
Db 188044 GAGGCTGACGTAACATTTAAAGTGGTAAAGAAATTTAATAACAGTATCAGAACGCGCA 188103

```

QY	181	TTAGGTTCCGATGTGTAATGCAATCAATTAACACACAGGCGAACAAGTTATTAAATAGTTCAA	240
Db	188104	GTTCGCGCGACGCTAATGAAAGCGCTTAACACCGGCTCAACAAGTAATTAATAATCGTTCAA	188163
QY	241	GATGAATTAACGAAGTGTATGGTGGAGAAAATACATCGATTATATGTGTCRAATAAACCA	300
Db	188164	GAAGAAATTAACAGCCTCATGGCGCGCAAGAAAGTAATAATCGGAACAGCTGATCGCCGG	188223
QY	301	CTTACTGTTTATGATGTTGGTTTACAGGTGCTGTTAAACACAACTCCAGGTAAA	360
Db	188224	CCAACCGTTATATGATGTTAGGTTTACAGGAGCTGCTAAACCAACCACTTCAGGAAA	188283
QY	361	TTAGCATATATGTCGTTAAATAATACACAAAACCAATCTATGTTGCGACAGATATT	420
Db	188284	CTCGCTAATTATTTACGTTAAATAATATACTGTAACCTTTACTTACTGTCGACACAGATT	188343
QY	421	TATCGTCCAGCAGCGATTAATCAATTCAACAAGTAGGGAACAATAATGATATTCCTGTGA	480
Db	188344	TACCGNCTGCACCATCAACATTTAGAAACACACTTGCACAACTATTAGATGCCAGTA	188403
QY	481	TACAGTGAAGSAGATCAAGTAAAGCCACAAATAATGTAACATATGCACTTAAACATGCT	540
Db	188404	TTTTCTCTAGCGCATCAAGTAAGCCAGTAGTAATTCGCGAACAAGTATCGCTAAAGCA	188463
QY	541	AAAGAAACATTTAGACTTTGTAATCATTCATGATACAGCAGCTCGATTACATCGATGAA	600
Db	188464	AAAGAAACATTTAGATTATGTGTCATTATCGATCAGCTGGTGCTTCTCATATCGAGAA	188523
QY	601	GCATTGATGACGAAATTAAGAAAGTAAAGAAATTTGCTAAACCAACCAAAATATGTTA	660
Db	188524	ACTCTGATGACGAATTAACACAGTCGAAGAAATCGCTACGCCACTGAATTTTACTT	188583
QY	661	GTTCGTGATTCATGACGGGTCAAGTCTGTCTCATGTTGCGAATCTTTTGACGATCAA	720
Db	188584	GTAGTGTATTCATGACTGGCGACAGCGCAAGTAATTTGGCCCAAGCTTCAACGAACAA	188643
QY	721	CTTGATGTCACAGCTGTACCTTAACATAAATAGATGTGATACAGTGTGTCGACGT	780
Db	188644	TTAGAAATACCGCGCTGTATTACAAATATAGCGGTGATACAGGTGGGCGACGA	188703
QY	781	TTATCTATTCTGTTGCGTGACAAAAACCAATTAATTTGGTATGATGATGAAAAAGTTA	840
Db	188704	CTTTCCATCCGTTCACTCACAGGAACAACATCAATCGTTGCTACCGGTGAAAAATG	188763
QY	841	GATGTTTAGCTATTCACATCTGAACGTATGCGATCACTCATGATTTAGTATGGGTGAT	900
Db	188764	GAAGCAATCGAAACGTTCCATCCGATGATGGCTTCAAGAATTTCTCGCGATGGGTGAT	188823
QY	901	GTGTTTAAGTTTAATGAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA	960
Db	188824	GTACTTTCTTATTGAAAAGACCAACTGATGTAGATGCGAAAAAATGAAACCTATG	188883
QY	961	GAGAAAAAGATGGGTGAGTCATCGTTTACTTTAGATGATTTTTTAGAACAACTTGATCAG	1020
Db	188884	GAACAAAAAATGAAGACAAACAGCATCACTTCTTTGGACCAATTTGCAACAA	188943
QY	1021	GTGAAAAATCTAGGACCACTGGATGATTAATGAAAAATGATTCAGTATGATTAATAATG	1080
Db	188944	GTAAAAAATATGGGACCACTAGATGAATCTACTAAAAATGATCGACGGGCAACAAATG	189003
QY	1081	AAAGGGCTAGATAAGCTTAATATGAGTGAAGACAAATGATCATATTAAAGCAATTATC	1140
Db	189004	AAAGGCTTCGACACATCAAGTAGTATGAATAACACTCGGTGCATCGACGCGATATT	189063
QY	1141	CAGTCATATGACCGCGCTGAAGAAAAACAATCCACAGACATTTGAATGTATCAGTAAAAAG	1200
Db	189064	AAATCCATGACAAAAACGAAAAAGATAATCCGGACATCATCAATCGACGAGAGAAA	189123
QY	1201	CGTATTCGTAAGGGTCTGGTCGTTTCAATTACAGAAAGTCAATTCGTTTGATGAAACAFTT	1260
Db	189124	CGAATTTGCTGGAGACGGACGCCAATTTACAGAAATCAATCCGCTCTTAACAACAFTT	189183

Qy	1261	ACGATATAGARAAATGAGRAACAATTCTACTGGTGGCGGTAAAGTTAAAAAAGGTAAA	1320
Ds	189184	GCTGAATGATAAAAAATGATGAGCAAAATGACTGTGGAGGAAAAGGTAAAGAAGGTAAA	189243
 RESULT 7			
AX417046/c	AX417046	349980 bp	DNA linear PAT 14-JUN-2002
LOCUS	Sequence 4037 from Patent WO228991.		
DEFINITION	AX417046		
ACCESSION	AX417046.1	GI:21449656	
VERSION			
KEYWORDS			
SOURCE	Listeria innocua.		
ORGANISM	Listeria innocua.		
REFERENCE	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.		
AUTHORS	I Glaser, P. and Kunst, F.		
TITLE	Listeria innocua, genome and applications		
JOURNAL	Patent: WO 022891-A 4037 11-APR-2002;		
FEATURES	Pasteur Institut (FR)		
	Location/Qualifiers		
	1..349980		
source	/organism="Listeria innocua"		
	/db_xref="taxon:1642"		
	by-seq 2058; original length: 3.011.208 replaced		
	/note=seq 2058; 0.000.001 to 0.349.980-seq 4032; 0.300.001 to		
	0.649.980-seq 4033; 0.600.001 to 0.949.980-seq 4034;		
	0.900.001 to 1.249.980-seq 4035; 1.200.001 to		
	1.549.980-seq 4036; 1.500.001 to 1.849.980-seq 4037;		
	1.800.001 to 2.149.980-seq 4038; 2.100.001 to		
	2.449.980-seq 4039; 2.400.001 to 2.749.980-seq 4040;		
	2.700.001 to 3.049.980-seq 4041; 3.000.001 to 3.011.208"		
BASE COUNT	101055 a 72969 c 50688 g 115268 t		
ORIGIN			
Query Match	51.2%; Score 700.8; DB 6; Length 349980;		
Best Local Similarity	70.7%; Pred. No. 9.8e+109;		
Matches 933; Conservative	0; Mismatches 387; Indels 0; Gaps 0;		
Qy	1	ATGGCATTGTAAGCGTTATCAGAACGCCCTGCAGCAGCATGCAAAAAATCGTGSTTAAG	60
Ds	129501	ATGGCATTGTAAGGACTAGCTGGAGACTCCAGAACACATGAACTTCGGCGTAAA	129442
Qy	61	GGTAACTTACTGAAGCTGATATAAAGATAATGATCGGTGAAGTAAAGATTAGCGTTATTT	120
Ds	129441	GGGAAAGTAAACGAAGCTGACGTAAAGAAATGATCGGTGAAGTTCGTCTGCTTACTA	129382
Qy	121	GAGCGTGCAGTAAACTTTAAAGTGGTAAAGCAATTTATTAACACAGTATCAGAACGGCA	180
Ds	129381	GAGCCGATGTTAATTAAAGTCGTTAAACAATTTATTAACACAGTAGCACACGTGCA	129322
Qy	181	TTAGGTCCGATGTAATGCAATCATTAACACCAGGCGCAACAAGTTATTAATAATAGTTCAA	240
Ds	129321	GTTGGCGGGACGTAATGAAAGCCGTACACCCCGGTCAACAAGTANTTAAAAATCGTTCAA	129262
Qy	241	GRNGAATTAACGAGTTGATGGGTGGAGAAATACATCGNTAATATGTCAAATAAACCA	300
Ds	129261	GAAGATTAACACGCCATCATGGGGGGGAGAAAGTAAATCGAACACAGCTGATCGCGCG	129202
Qy	301	CCTACTGTGTTATCATGPTGTTTTACAAGTGCTGGTAAAAACAACAACTGCAGGTAAA	360
Ds	129201	CCAACGGTTATTATGATGTAGTTTACAAGSAGCTGTTAAACAACACCACCTTCAGGAAA	129142
Qy	361	TTAGCATTTATTCATGCGTAAAAAATACACAAAAACCTATGTTAGTTCGACAGATATT	420
Ds	129141	CTCGCTAATTTTATTCGTAAAAAATATATCGTAAACCTTTACTAGTCGCACAGATATT	129082
Qy	421	TATCGTCCAGCAGGATTAATCAATTACAACAGTAGGGGAAACAAATGATATCTCTGTA	480
Ds	129081	TACCGNCTTCGCGCAATCAACAAATTAGAACACTTGGCAACAATTTAGATATGCCAGTA	129022
Qy	481	TACATGTAAGGAGATCAAGTAAAGCCACAACAAATTTGTAACATTAATGCAATTAACATGCT	540

Db 129021 TTTTCTCTAGCGATCAAGTAAAGCCAGTAGAAAATCGGCGAACAGAGCTATCGCTAAAGCA 128962
QY 541 AAAGAAACAACTTTAGCTTTTGTATCATGATACAGCAGGTGCTATACACATGATGAA 600
Db 128961 AAAGAAACAACTTTAGCTTTTGTATCATGATACAGCAGGTGCTATACACATGAA 128902
QY 601 GCATTGATGAGCAATTTAAAGAGTAAGAAATGCTAAACCAACGAATATATGTTA 660
Db 128901 ACTCTGATGAGCAATTTAAAGAGTAAGAAATGCTAAACCAACGAATATATGTTA 128842
QY 661 GTTGCTGATCAATGACGGGTCAAGTGTCTCAATGTGTGAGAAATCTTTTGTGCGATCAA 720
Db 128841 GTAGTTGATTCATGCTGGCGAGACGCGAGTAATGTGGCCCAAGCTTCACAGAACAA 128782
QY 721 CTTGATGTCACAGGTGTTACCTTAACATAATTAGATGCTGATACACAGTGTGCTGAGCT 780
Db 128781 TTGAATAATACCGGCTTTGTTATTAACAAATTAGACGGTGTATACAGTGTGGCGAGCA 128722
QY 781 TTATCTATTCGTTGCTGTCACACAAACCAATTAATTTGTTGTTGATGAGTGAAGATTTA 840
Db 128721 CTTTCCATCGGTTCAGTCACAGGAAACCAATCAATTCGTTGCTACCGGTGAAATAATG 128662
QY 841 GATGGTTTATGAGCTATCCATCTGAACGTATGCGATCAAGTATTTAGTATGAGTGTGAT 900
Db 128661 GAGCAATCGAAACGCTTCCATCCGATCGTGTGCTGCAAAATTTCTCGGCATGGTGTAT 128602
QY 901 GTGTTAAGTTAATTGAAAGAGCGCAACAGATGTGGATGAGAAAGCAAAAGATTTA 960
Db 128601 GTACTTCTCTTATGAAAGAGCAAACTGATGTAGATGAGAAATAATGAAGCTATG 128542
QY 961 GAGAAAGATGCGTGTGCTATCTTTTACTTTAGATGATTTTGTAGTATGATGATGATCAG 1020
Db 128541 GAAACAAATGAAAGAGCAACAGCATGACCTTAGATGACTTCTTGGACCAATTCGAACAA 128482
QY 1021 GTCAAAATCTAGACCTAGCTGATGATTTATGAAATGATTCAGATGATGAATAAATG 1080
Db 128481 GTAAACAAATGGACCTAGATGATGAAATGATTCAGATGATGAATAAATG 128422
QY 1081 AAGGGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
Db 128421 AAGGGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 128362
QY 1141 CAGTCAATGAGCGGCTGAAAGAAACCAATCCAGACACATGATGATGATGATGATGATG 1200
Db 128361 AATCCATGACCAAAAGAAAGATTAATCCGACATCAATGATGATGATGATGATGATG 128302
QY 1201 CGTATGCTAAGGGTCTGGTCTGCTATTAAGAGATGATGATGATGATGATGATGATG 1260
Db 128301 CGAATTGCTGCTGAGAGCGGCGGCAATTCAGAAATCAATGCTGCTGCTGCTGCTG 128242
QY 1261 AACGATATGAAGAAATGATGAACAAATTCAGTGTGGGCTGATGATGATGATGATGAT 1320
Db 128241 GCTGAATGAAATAATGATGAGCAAAATGCTGTTGGGGAAGAGGTAAAGGATGA 128182

RESULT 8
LOCUS AX416333/c
DEFINITION Sequence 3324 from Patent WO228891.
ACCESSION AX416333
VERSION AX416333.1 GI:21448790
KEYWORDS
SOURCE Listeria monocytogenes ATCC 19115.
ORGANISM Listeria monocytogenes ATCC 19115
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1. Glaser, P. and Kunst, F.
AUTHORS Listeria innocua, genome and applications
TITLE Patent: WO 0228891-A 3324 11-APR-2002;
JOURNAL Pasteur Institut (FR)
FEATURES
Location/Qualifiers
i. .2256
/organism="Listeria monocytogenes ATCC 19115"

/db_xref="taxon:176281"
BASE COUNT 533 a 471 c 420 g 832 t
ORIGIN
Query Match 51.0%; Score 697.6; DB 6; Length 2256;
Best Local Similarity 70.5%; Pred. No. 1.2e-107;
Matches 931; Conservative 0; Mismatches 389; Indels 0; Gaps 0;
QY 1 ATGGCATTTGAGGCTTATCAGAACCCCTGCAAGCGACGATGCAAAAATGCGTGGTAAG 60
Db 1666 ATGGCATTTGAGGCTTATCAGAACCCCTGCAAGCGACGATGCAAAAATGCGGCGAA 1607
QY 61 GGTAACTTACTGAGCTGATATAAAGATATATGATGCGTGAAGTAAAGATATGCGTTATT 120
Db 1506 GCGAAATTAACGAGCTGACGTAAAGAAATGATGCGTGAAGTTCCTTCTCTACTT 1547
QY 121 GAGGCTGACGTAAACTTTAAAGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180
Db 1546 GAAGCGATGTTAACTTTAAAGTCGTTAAACAAATTTATTAACGCTGAAGCGATGCT 1487
QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACCGGCGCAACAAGTTTATTAATAAGTCTCAA 240
Db 1486 GTGCGCGGAGCTTATGAAGAGCTTAACACCCGCGCAACAAGTTTATCAAAATCGTTCAA 1427
QY 241 GATGAATTAACGAAGTTCATGCGTGGAGAAATATCATGATGATTAATATGTCAAATTAACCA 300
Db 1426 GAAGACTTAACGAAGTTCATGCGTGGAGAAATTAACCAATCGGAGCGGCGCGCA 1367
QY 301 CTTACTGTTGTTATGATGTTGTTTACAAAGTGTGTTTACAAAGGCTGGTAAACACAGCTTCAA 360
Db 1366 CCAACCGGATGATGATGTTGTTTACAAAGGCTGGTAAACACAGCTTCAA 1307
QY 361 TTAGCATTTATGATGCTGATAAATAATACAAACCAACCTATGTTAGTTCAGCAGATATT 420
Db 1306 CTTGCTTAATTTATGCAAAATAATATAACCGTAACCACTTCTGTCAGCAGATATT 1247
QY 421 TATCTCCAGCAGCGATTAATCAATTAACAGTAGGTAAGCAAAATGATGATATTCCTGTA 480
Db 1246 TACCGACCTGCGGCAATCAACATTAAGAACACTTGGCAAGCAATTAATATGCGGTA 1187
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCCAACAAATTTGTAACATAATGCAATTAACATGCT 540
Db 1186 TTTTCTCTAGGACCAAGTAAGCCCGGTAGAAATCGCCAAACAGCTATGCTAAAGCA 1127
QY 541 AAAGAGACATTTAGACTTTGTAATCATGATACAGCAGGTGATTAACATGATGATGAA 600
Db 1126 AAAGAGACATTTAGACTTTGTAATCATGATACAGCAGGTGATTAACATGATGAA 1067
QY 601 GCATTGATGAGCAATTTAAAGAGTAAGAAATTTGCTAAACCAACGAATTTATGTTA 660
Db 1066 ACCTTAATGATGATTAACAGCTGAAGAAATTTGCAACGCGCAACCGAAATCTGCTT 1007
QY 661 GTTGTGATTAAGCAGGCTGAGTGTGCTCAATGTTGAGAAATCTTTGACGATCAA 720
Db 1006 GTTGTGATTAAGCAGGCTGAGTGTGCTCAATGTTGAGAAATCTTTGACGATCAA 947
QY 721 CTTGATGTCACAGGTTTACCTTAACCTAATTAATGATGATGATGATGATGATGATGATG 780
Db 946 TTAGAAATTTAGAGTGTGCTGCTAACCATAATGATGATGATGATGATGATGATGATG 887
QY 781 TTATCTATTCGTTGCTGACACAAACCAATTAATTTGTTGATGATGATGATGATGATG 840
Db 886 CTTTGGATTCGCTGCTGACGCAACCAATTAATTTGCTGCGACTGCTGCTGCTGCTGCT 827
QY 841 GATGTTTATGAGCTATTCATCTGACCTGATGATGATGATGATGATGATGATGATGATG 900
Db 826 GAAGCTCTAGAAATCTTCATCCAGATGATGATGATGATGATGATGATGATGATGATG 767
QY 901 GTCTTAAGTTTAAATGAAGAGCGCAACAGATGTTGATCAAGAAAGCAACAAAGATTTA 960
Db 766 GTCTTCCCTGATGAAAGAGCGCAACCGATGATGATGATGATGATGATGATGATGATG 707
QY 961 GAGAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020

```

Df 706 GAACAAAAATGAAGAACACACAGCATGACGCTAGATGACTTCTAGACAGTTACAACAA 647
QY 1021 GTGAATAATCTAGACCACCTGATGATATATGAAATGATTCACGATGATGAATAAATG 1080
Df 646 GTAAACAAATGGGACCATAGATGACTTAAATGATGCGAGGCGCAACAAATG 587
QY 1081 AAAGGCTAGATAAGCTTAAATGATGAGTGAAGCAAAATTCATCATATTAAGCGATTATC 1140
Df 586 AAAGGCTCGACACATGACGTCGATGATGAACAACTAGTCAATGAGCAATATC 527
QY 1141 CAGTCAATGACGCGCTGAAGAAACAATCCAGACATTTGAATGATGATGATGATGATG 1200
Df 526 AAGTCAATGACCAAAATGAAGAAACAATCCAGATATCATCAATCAAGTAGAAGAAA 467
QY 1201 CGATGCTGAAGGCTGCTGCTTCATTAACAAGATCAATGATGATGATGATGATGATG 1260
Df 466 CGATGCTGACGCTGACGAGCTCGGTTCAAGAAATTAACCGTCTCTTAAGCAATTT 407
QY 1261 AAAGGCTAGAAATGATGAACAAATTCATGCTGCGGTAAAGTGAAGTGAAGTGAAGTGA 1320
Df 406 GGTGAATGAAGAAATGATGAAGCAATGATGATGATGATGATGATGATGATGATGATG 347

RESULT 9
Df4356
LOCUS
DEFINITION Bacillus subtilis orf1, f1f, rps genes for ORF1, f1f and 305
riboseomal protein S16, complete cds.
ACCESSION Df4356
VERSION Df4356.1 Gi:439700
KEYWORDS rps; orf1; 305 ribosomal protein S16; f1f.
SOURCE Bacillus subtilis (strain.168) DNA.
ORGANISM
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (bases 1 to 2739)
Honda,K., Nakamura,K., Nishiguchi,M. and Yamane,K.
Cloning and characterization of a Bacillus subtilis gene encoding a
homolog of the 54-kilodalton subunit of mammalian signal
recognition particle and Escherichia coli F1f
J. Bacteriol. 175 (15), 4885-4894 (1993)
93328695
REFERENCE
2 (bases 1 to 2739)
Yamane,K.
Direct Submission
Submitted (03-FEB-1993) Kunio Yamane, Inst. Biological Sciences,
Tsukuba University, Tsukuba, Ibaraki 305, Japan (Tel:0298-53-6419,
Fax:0298-53-6006)
FEATURES
Location/Qualifiers
1..2739
/organism="Bacillus subtilis"
/db_xref="taxon:1423"
292..297
/evidence=experimental
315..320
/evidence=experimental
365..697
/gene="orf1"
365..697
/gene="orf1"
/codon_start=1
/transl_table=11
/product="ORF1"
/db_xref="GI:242968"
/translation="MSLEKTRNMLDFYQSLTSSKSYMSLYLDDFSLGEIAEE
YEVSRQAVVDNIKRTAEMLEQYEEKLLLLKFKERKFNKELASGSGKEEITAL
ITALKELD"
711..2051
/gene="f1f"
711..2051
/gene="f1f"

```

```

/codon_start=1
/transl_table=11
/product="F1f"
/protein_id="BAA21691.1"
/db_xref="GI:2309080"
/translation="MATEGLADRLQOTISKIRGKGVSEQDVKEMREVRLLLEADY
NFVKYKDFVKYSEFAYGVODVMSLTPGOVLYKVQBELTLMGEBESKTAFAKRPPT
VIMVGLQAGKTTTSKLANLKKHKNKPMVAADIRPAARINQLETIGKOLNMPV
FSLGDDVSEVETAKOAEKESEHYDIVILDTAGRLHIDHDELMDELNVKEIANPEEI
FLVDSMTGQDANNVAKSFNEQLGLTGVVLTLDGDTGRGAALSIRAVTNTPIKFAGL
GERLDALEPFHMASRIILMGVDLTLLIEKAQASVDEDEKAELEQKMRMTSLTDDF
LEOLGVRNMGPLDELLQMPGAKGNLKNIOVDKOLNHNVEAIKSMVLEKEQPD
IINASRRKRIAGSGTSTVQEVNRLKQFDEMKNMKQMNTNMSKCKKXGKFLPW"
-35_signal 2091..2096
-10_signal 2117..2122
gene 2157..2429
/genes="rps"
2157..2429
CDS
/genes="rps"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S16"
/protein_id="BAA21692.1"
/db_xref="GI:2309081"
/translation="MAVKIRLKRMAKSKSPFYRVIVADSRSPDRGFIETVGTYNPVA
KPAEYKIDEEALALWLOTGAKPSDTYRNLFSSQGIEMXFNHAKGK"
BASE COUNT 887 a 527 c 648 g 677 t
ORIGIN
Query Match 45.5%; Score 622.8; DB 1; Length 2739;
Best Local Similarity 67.1%; Pred. No. 3.9e-95;
Matches 882; Conservative 0; Mismatches 432; Indels 0; Gaps 0;
QY 1 ATGGCATTGGAAGCTTATCAGAACGCTGCAACGACGACGATGCAAAAATGCGTGTAAAG 60
Df 711 ATGGCATTGGAAGGATTAGCGACCGACTGCGACGACGACGATTTCTAAATTCGCGGAAA 770
QY 61 GGTAACTTACTGAAGCTGATTAAGATATATGATGCGTGAAGTAAAGTATGCGTATTT 120
Df 771 GGGAAAGTCAGCAACAAGATGTAAGAGATGATGATGCGTGAAGTATGCGTGTGT 830
QY 121 GAGGCTGACGTAAATTTAAAGTGTAAAGAAATTTATTAAGACATATCAGAACGCGCA 180
Df 831 GAGGCTGACGTAACTTTAAAGTGTAAAGATTTTGTTCAAAAGTAAAGTGAAGCGCT 890
QY 181 TTAGTTCGCGATGATGCAATCATTAACACGCGGCAACAACTTATTAATAATAGTCAA 240
Df 891 GTAGSCCAAGACGTCATGAAAGTCTGACGCGCGCCGACGAGCTCATTAAGATGTTCAA 950
QY 241 GATGAATTAAGACGATGATGCGTGAAGAAATATCATCGATTAATATGTCAAATAACCA 300
Df 951 GAGGACGTGACTGAGCTGATGGGCGGCGGAGAGAGCAAAATCCCGTCGCAAAAAGCGG 1010
QY 301 CTTACTGTTGTATGATGTTGTTTACAAGTGTCTGTTTAAACAACTGAGTAAAGTAA 360
Df 1011 CGGACTGTTATGATGTTGTTTCCAAGTGTCTCCAAGTGTCTGTTTAAACGACGAGTAA 1070
QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAACTATGTTAGTTGCGACGATATT 420
Df 1071 CTTGCAATCTGCTGCGCAAAAAGCATTAATGCAAAACCGCATGCTGTTGCTGCGATATT 1130
QY 421 TATCGTCCAGCAGGATTAATCAATTAACACATAGGGAACAATGATATTCCTGTA 480
Df 1131 TACGCGCCCGCCCAATTAAGCATCTGGAACACTCGGCAACACGCTGATATGCTGTT 1190
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATGTCATTAAGTAAACATGCT 540
Df 1191 TTCCTCTCTGCGATCAGTCACTCTCTAGAAATAGCTTAACAGGCTATTGCAAGAGCC 1250
QY 541 AAGAGAACATTTAGACATTTGTAATCATTTGATACAGCAGGTCGATACATCGATGAA 600
Df 1251 AAGGAGAACATTTAGCTACGTCATTTTGGATACGCGCGCTTGCATATCGACCAT 1310

```

Klaer-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
Koningsstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
Larindois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
Masuda,S., Maueil,C., Medique,C., Medina,N., Mellado,R.P.,
Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
Ogawa,K., Ogihara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
Portetlelle,D., Porwollik,S., Prescott,A.M., Prescean,E., Pujic,P.,
Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadafire,Y., Sato,T.,
Scanlan,E., Schleich,S., Schroeter,R., Scodone,F., Sekiguchi,J.,
Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstr,P., Tognoni,A.,
Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenecker,T.,
Winters,P., Wipat,A., Yamamoto,H., Yamanue,K., Yasumoto,K., Yata,K.,
Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
Dauchin,A.

The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377

2 (bases 1 to 208780)
Kunst,F., Ogasawara,N., Yoshikawa,H. and Dauchin,A.
Direct Submission
Submitted (18-NOV-1997) I. Mosser, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: mosser@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

Location/Qualifiers
1. 208780
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
258. 4588
/gene="bpr"
258. 4559
/gene="bpr"
/EC_number="3.4.21.-"
/note="alternate gene name: bpf"
/codon_start=1
/transl_table=11
/product="bacillopeptidase F"
/protein_id="CAB13404.1"
/db_xref="GI:2633903"
/db_xref="SWISS-PROT:P16397"
/translation="MRKTKNRLISLVSTVVVLSLFPGAAGASKVTSVVKELQ
SASQNTKISSLRKSKKKKEKTKFLIFKDLNPEKAAAVKAKSKLSAAKTY
QKRSVASLKVTADESQDVLNLTQDKGNADQIHSTVYVNGIAVHASKEYMYR
VQFVEYKFLNKRQQLFKSSSPFNKKAKRAIKATQGVENVNDQIDAPKAWALGYDG
THTVMSIDTGVENVNHEALKEKGYFNPEENENENMNDVAGEASPYDILAHGT
HVTGVSQSGDPGNTQIVGAPGAKITFNKASQEGDQDAILPAGWILPDKAEGNP
HPMAPQVNNKSGGSLDLEWRDMYNARARADLPFSAGNTDLETPGGPGSTANP
ANYPSPATGATDINKLADSLQGPSDEIKETLSAPGVNIRSSVPGQ2YEGDGNV
TSAGHPSVAVALKQANSLQSVSEMEDILITAEPLTUSTPSTPGNNGHGLWNA
FDVASVGLGUGKAEGVSGVEDGGPPYQHEKVIYAGTGSLSLPTLTAEDNYSVTS
KLSVTKDQSGWETTKRISLDHGLQYAEIPDKTGLSKYKMIHDFGHWVSSD
VIDVTYVSTAGTKGTFKYVDTPGGWVAGSTNNWNGVTPGNTAPASGEKVYGLT
GNYANSANNLVMPPIKAPDSGSLFQKPSWHNLEDYGYVFLVPGENRQAGV
YNGKTSWTDSEIDLSAYGQNTQWNLQSDSIAKEGWYIDVVLSPKSAGTKYK
NLQVGEPSGQKLPYNNKAKKPSANTAVYHQNKAIQEQVLPKQAVSVVETGKSTY
SQOSTGQTLKHKAGDTILMAEYAGQSKQKSLVTDQANFTLEMKKGTGKGT
VINTKGTEPVGASVYVEDAVEPAMTNDKGETMLEYEGATYTKVAPGYSDFS
KLVGDKETALPQFVYGPGETIADGTANANSYFAGNGWAVKMTLDGKDGML
TGGLRFWMTDFPDGGTEFKYVDGKQPGKKIAGPFAEALRNGEWTVGYMI
SKGIMVDDQFLVLYIQSKPDPYSPGLANDETGQNSGNQYIDGKWQDGDAGVMI
RALYDEAAVPEIISPDSYNNKQSVYVYKNASPGTTHYNGEKEAGETKAAADGT
FHAGLILNGNEUTAFSTNGTIDASPSPITVLDQKPEPLTDNPDDGKTKNETI
TVKGVASDNLKQVKNKAGSVADSPISAEIENRNEIKVITDLAGNKTTKTY
IDVNFDPFVLSGLIPGBDKNLKAGSVIAFSSAIEDADITFIRLPTLNARASVQNT
ELP.LAEIPSGYEGWYATSSIKAGKAVYEVSDGYGNETRKTANGKLNNTEN"

```
terminator
4571..4588
/feature="bpr"
gene
4754..5683
/feature="spoilGA"
CDS
4754..5683
/feature="spoilGA"
/function="processing of pro-sigma-E (SpoIIGB) to active sigma-E (stage II sporulation)"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="CAB13405.1"
/db_xref="GI:2633904"
/db_xref="SWISS-PROT:P13801"
/translation="MKIYLDVILWLNFCDFALLLLAFILKRVKRLVGGAFIGS
SVLLMPTFPPPIVEHPAGKIAFSVIVVVFGRFRFFQNLFSFVATELGGGI
IGAHLSGINSIQVGMVINGTGDPISLWLFVGGFPAWFSKRIEDIEKNQI
YERVSQADLGSTLHVRLGIDSGNQLIDPLTPWMIYIDKLEPIFGTAETMIR
NTPLEALTEQDDSFRLDKMLRILPYRGVQGNQFLLCVKPDHVTIMTKEEMISADKC
LIGISTKLSDAGEFDALIHPRKMLSGKAVKVS"
5746..6465
/feature="sigE"
CDS
5746..6465
/feature="sigE"
/function="early mother cell-specific gene expression"
/note="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29 factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKKLRLTLWYKLMKLGKSDVEYVIGSGEALPPPLKDEE
QVLLMKLPNGQARALLERILRLVIRKFNENGINEDLSIGTIGLIKAVNPF
NPKIKLMTVAGSCNIEMLMLRNKRIKSEVFEPLNIDWDGNELLSDVLGTD
DDITKIDLEANVDKLIKALKLEQLNREKQINELRFLVGLVEEEKTKQVADMIGISOS
YISREKRIIKRLKKEFNKV"
6605..7387
/feature="sigG"
CDS
6605..7387
/feature="sigG"
/function="late forespore-specific gene expression"
/note="alternate gene name: spoIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor (sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSRNKVEICGVDTISKLPVLKNEEMRKLFRLQDEGDSAREKLV
NGNRLVLSVIOFRNNGEYVDDLPQVCGICGLMKSIDNFDLSHWKFSYVAPMIGE
TERVLBNRIRVRSRLRIAYKALQVRERLISSEKPEADIAKLVFPHEEIVFA
LKAIDPFQKTEPIYNDGDPYIVMDQISDERNTDSOWIEELAKVLRGMRUNDRKMI
LRKIDPFQKTEPIYNDGDPYIVMDQISDERNTDSOWIEELAKVLRGMRUNDRKMI
7535..8496
/feature="ylmA"
CDS
7535..8496
/feature="ylmA"
/function="unknown"
/note="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPREMBL:O31723"
/translation="MIQLDNVSLKRNKGWILKDIHWKVEKENWVYGLNGAGKIAL
NMLGSIYFTPSGEMRVIGHEFGKTELGRKLRKIGLVSAALOOKLYPADSAFIALS
GAYAGISYFTPSGEMRVIRKAIKGLLEDGAIEYADRRYETLSQGEKORALIRALMADP
ELLILDPEVGLDFAREKLITITYIANKENAPSILYVTHAEELIPYFDKALLLKQ
GVYFSGSIEKMLTDQLLSAFFDPIHLVNWQDRPFLTRAEPITNA"
8472..8496
terminator
```

```

/feature="ylmA"
8531..9811
/feature="ylmB"
8531..9811
/feature="ylmB"
/function="unknown"
/note="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11
/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPREMBL:O31724"
/translation="MDQIYSIQKKVEHKEELIQLAKTLISYQTAPPARNTGSIQS
WHPPIPKNGLLIGRGASDMKGMACVLFVKLIRASIELPDILLOSVIGEEVGE
ASTLECCRGVHADPAIVADTSDMHIOQGGVITGWIEIKSSOTFHDGRTRRNIRAGG
GTFGSALERKAKIAGLELKEHWSIMKSPGFKPTGNTINFAVEGGRHAAFTADE
CLMTLVHFFYNETHQVRAEIEDVYNLSDSDIWLRENRPVFKWGGSSMIIDRGEIF
PALEYDPGPGVLAATASHOKVRCFCPIIDVSQSVTDGWLVDAGIPCVIYGGDLHN
AHSVNEKYSIQLVVEYTKIILDFIISWCSRKEQ"
9894..10139
/feature="ylmC"
CDS
9894..10139
/feature="ylmC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPREMBL:O31725"
/translation="MISISBFQKDVNVNNGKLGISGIDIDINVTGKIQAILGGN
Query Match 45.5%; Score 622.8; DB 1; Length 208780;
Best Local Similarity 67.1%; Pred. No. 1.3e-95;
Matches 882; Conservative 0; Mismatches 432; Indels 0; Gaps 0;
QY 1 ATGGCATTTCAGCGCTTACAGACGCGCTCGACGACGATGCAAAAAATCGGTGGTAAG 60
DB 73153 ATGGCATTTCAGGATTACCGCGGACTGCGACGACAGCATTCCTAATAATCCGCGGAAA 73212
QY 61 GGTAACTTACTGAAGCTGATATAAAGATATGATCGGTGAAGTAAAGATTAGCGTTATT 120
DB 73213 GGGAACTGACGCAACAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 73272
QY 121 GAGCTGACGCTAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGCGCA 180
DB 73273 GAGCTGACGCTTAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGCGCT 73332
QY 181 TTAGGTTCCGATGATTAATGCAATCATTAACACGAGGCAACAAGTTATTAAATAATAGTTCAA 240
DB 73333 GTAGCCCAAGACGCTCATGAAAGTCTGACGCGCGCCGACGAGTCTAATAGTTGTTCAA 73392
QY 241 GATGAATTAACGAAGTGTGATGGGTGGAGAAAATACATCGATTAATATGTCATAAATAACCA 300
DB 73393 GAGGAACCTGACTGAGTGTGGCGGGGAGAGAGCAAAATCGCGCTGCGCAAAAGCGCG 73452
QY 301 CCTACTCTGTTATGATGTTGTTTCAAGGTCGTGTTAAACACACAACTCCAGGTAAA 360
DB 73453 CCGACTGTTTATGATGTTGTTTCAAGGTCGTGTTAAACACACAACTCCAGGTAAA 73512
QY 361 TTAGCATTTATGATGTTAAATAATACAAACAAAACCTATTGTTAGTTGCACAGATATT 420
DB 73513 CTTGGAATCTGTCGCAAAAAGCAATATCGCAACCGATGCTGTTGCTGCGGATATT 73572
QY 421 TATCGTCCACAGCATTAATCAATTAACAACAGTAGGAGAAAACAATATGATATTCCTGTA 480
DB 73573 TACCGCCACCGCAATTAAGCAGGTGGAACACACTCGGCAAAACAGCTTGTATATGCTGTT 73632
QY 481 TACAGTGAAGAGATCAAGTAAAGCCACCAATTTGTAACCTAATGCATTAAACATGCT 540
DB 73633 TTCTCTCTTGGGATCAGGTCACTGCTGTAGAAATAGCTAAACAGGCTATTGAGAAAGCC 73692
```



```

/translacion="MSLKNQIKSEHDHNLKLVSONVTIVFALMLFFAILEFIFGV
LRFNSAIVLSADVHTGDVAIGLAWFQKFSKKREDKHSFGQORSLLGASLTGVL
ITGCSFVLEAVREFNPQPVATGMGLAIFAIYANGFGAWLLARGSSNESILNLH
ALEDVHGLVLIYSIVYHFVXWLDPLLSILIALFSLKAPKFWGLRILLESYP
EDDKLRALQELQEPVLAHTVOLIWSIDGQNAAMIHIVIPENQDFSDAKIAVRKL
LESEKVSASAELEDDETHKHHQVEI"
gene      complement(2940. .4181)
          /gene="proa"
          /note="I0116"
          /complement(2940. .4181)
          /gene="proa"
          /note="EVIDENCE BY HOMOLGY BIO01.05 Glutamate family.
SECOND STEP IN PROLINE BIOSYNTHESIS PATHWAY"
          /codon_start=1
          /transl_table=11
          /product="gamma-glutamyl phosphate reductase (EC
1.2.1.41)"
          /protein_id="AAK05706.1"
          /db_xref="GI:12724615"
          /translacion="MIEELGLVKKASKEVAKLSTADKNVFLQNLADSLIENTDRIIS
ENEKDLANALEHGISETMDORLRLNAQRI SDMATGLROVAELPDPIROVLOGFTNLDG
IKTLQKRVPLGTVMGIVPESRPNTVIDAFSCFKTGNVLLRGSSDALYSNMVLELIK
ENLLSKATDGAVELLSDTSHAERKMQADKFDVLIPRGSRKLNRKVEKATVPVI
ETGVGNCITFVDESADLEMATKIVINAKTORPSCVNAESLVHAKIADIEFUPKLENE
INKVHEIFERADRALKVLGAGIPATDDDFGEFLDYILSVKYVDNLDEAIBHINTYS
SRHSISIVHDYFNAQKFPQDEIDAAAAYVYNASTRTDGFVEGLGAETIGISTQKLEARG
PMGLEALTSTKYLIDGTQIR"
gene      complement(4199. .5011)
          /gene="proB"
          /note="I0117"
          /complement(4199. .5011)
          /gene="proB"
          /note="EVIDENCE BY HOMOLGY BIO01.05 Glutamate family.
FIRST STEP IN PROLINE BIOSYNTHESIS PATHWAY"
          /codon_start=1
          /transl_table=11
          /product="glutamate 5-kinase (EC 2.7.2.11)"
          /protein_id="AAK05707.1"
          /db_xref="GI:12724616"
          /translacion="MIMTRKNILSVKRIYVVKIGTSSLLIPNGKINLSNIDELAFVLSO
LNKNGEVLIVTSAGVGLNVLGIDKRPKGIADQOALASIGQVELMSLTQMFRYS
QKVSQLLTRDVTDPYSRENAENALNALLGUGIPIINENDAIJVDMDHDPKTFGDN
DKLGAIVSKNADLIMLSDIDGLFKNPSYDDAKIFNEIHEITDELROMAGGAGS
RFTGGMSTKLAAQILFENGQWMLTNGRERIEIKETIEGREIGTFYFHQS"
gene      complement(5141. .6613)
          /gene="yqfA"
          /note="I52568"
          /complement(5141. .6613)
          /gene="yqfA"
          /note="EVIDENCE BY HOMOLGY BIO15.02 HYPOTHETICAL.
Conserved"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAK05708.1"
          /db_xref="GI:12724617"
          /translacion="MTNMNKSITIEROKRIVELITLLHEGGFMDAKKIFNEFDVGD
VSPITGAERALKOGGLNPSEIOLCNVHAAPKGAITDIHKSSLEENTPGHPHTMKL
ENQVITSLNDEIRVFSKIEKGDMYSKRELLAAVLDLNDKHAKESLIFSPNEK
IGTAPKVMVGVDSDVSKMEIKVIVLTKDTPALNPLEMLENLLTIEEMIFKEEA
IMTPICLDYFSLKOWEQIEEDSKSIEIGYFAEPALPKWAKSESIEKEHEPPQRLAE
SAKAMTEALAKESGLENSKPKRVKQYDWEKVTSEGVVLPFGMLHNLNLTALFNVL
VDSFMEKEDIVRRFSGGGERIFPRAKSVIGRVIDCHPPKSNVAVKILBDRKSGPRD
QADFWIDLHFNKITYIRYFAMRDEETGEVLGCLFVSODITISIQNLVGEKRLDGHKT
TSEYENQVAKSIKKSTDGAVNGENPDFVKMLAEKTKNS"
gene      complement(6613. .6852)
          /gene="yqfB"
          /note="I54040"
          /complement(6613. .6852)
          /gene="yqfB"
          /note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
          /codon_start=1
          /transl_table=11

```

```

/product="UNKNOWN PROTEIN"
/protein_id="AAK05709.1"
/db_xref="GI:12724618"
/translacion="MIQDINAKLYDLATEHPEIIDLMDDLGFHEIKMPGMLQTAGRM
ATIPGAKKHIDMDKIVIAFAPAGFEFSNQVEE"
gene      complement(6979. .7305)
          /gene="yqfC"
          /note="I54406"
          /complement(6979. .7305)
          /gene="yqfC"
          /note="EVIDENCE BY GENEMARK BIO16.01 Unknown"
          /codon_start=1
          /transl_table=11
          /product="UNKNOWN PROTEIN"
          /protein_id="AAK05710.1"
          /db_xref="GI:12724619"
          /translacion="MSSFESFISLTIGCAIVTWISRVLPFTLLKKMLPOLVVEVLSF
VPVIMSLAWISNLFIOHGLFSLVSNWNLASIPTVLAATLTKNLLVIVLVGVSFA
FIQIIP"
gene      complement(7292. .7999)
          /gene="yqfD"
          /note="I54719"
          /complement(7292. .7999)
          /gene="yqfD"
          /note="EVIDENCE BY HOMOLGY BIO13.02 TRANSPORT AND BINDING
PROTEINS. Amino acids, peptides and amines"
          /codon_start=1
          /transl_table=11
          /product="amino acid permease"
          /protein_id="AAK05711.1"
          /db_xref="GI:12724620"
          /translacion="MNSLFTFKQGLKDXMTPTVFGYIGIAFGMIGHSEGSFVWVILL
LSLIVYAGSAQPIVMSYATHSPIMSVLSVFLVNSRIILMSMTTASYPKNESLLKNI
LGLTLITDSEPALGMKNKONYTEGLKNSFNLSNLLAYLVWLASALGALLGNLANP
EKLGIGFAYIAFMFGLLYLQLISDKTLGLMLQLVMVGITLVFYFGLIFLPSNLLVLF
VTLIACALGVGVKRVFF"
Query Match      41.4%; Score 567; DB 1; Length 11023;
Best Local Similarity 64.7%; Pred. No. 6.3e-86;
Matches 843; Conservative 0; Mismatches 460; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGCCTTATCAGAACGCTGCAAGCAGCATGCAAAAATGCGTGGTAAG 60
Db 10954 ATGGCATTTGAAAATTAAGTGAACGCTTACAAAACGCTTTTAAAAATCTACGTGAAA 10895

QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAGAGNTAGCGTATT 120
Db 10894 AAAAAATTAAGTGAAGCTGATATAAGATAATGATGCGTGAAGTAGAGNTAGCGTATT 10835

QY 121 GAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAAACGCA 180
Db 10834 GAAGCCGAGTTGCTCTCCAGTTGTCAGAGAGTTTATTAAGGCAATCCGTGACAGCT 10775

QY 181 TTAGTTCGCGATGTAATGCAATTAACACAGGCAACAGTAATTAATAATAGTTCAA 240
Db 10774 GTCGCTGTTGAAGTATCAGAAAGCAATTAACCCAGCTCAACAGTAATTAATAATGTTGAT 10715

QY 241 GATGAATTAACGAAGTGTAGGCTGGAGAAAATACATCGATTAATATGTCAAATAAACA 300
Db 10714 GAAGAAATTAACAGCTAATTTGGGTGGTGGCAGCTGAATTAATTAATCTCTCTCTCT 10655

QY 301 CCTACTGTTGTTATGATGGTGGTTTACAAAGGTGCTGTAACAAACAAACACACAGTAAA 360
Db 10654 CCAACAAATTAATGATGGTGGTTGTCAGGCGCAGTAAACACACTTTTCTGCTGTA 10595

QY 361 TTAGCATTTATGATGCGTAAAAAATTAACAAACAAACAACTATGTTAGTTCAGCAGAT 420
Db 10594 CTTGCCAAAACCTCAAAAGAGAGCAAAATGCTCGCCCTTTAATGATTCAGCCAGCT 10535

QY 421 TATCTCCAGCAGCGATAATCAATCAATTAACACAGTAGGGAACAAATTTGATATCTCTGA 480
Db 10534 TATCTCCAGCAGCGATTGACCAATTTAAAAACCTTAGGTAGACAGCTTGAAATCTCTGC 10475

```


QY	421	TATCGTCGACGAGCGGTAATAATCAATACAAACAGTACAGGAAACAATATGATTCCTGTGA	480
Db	210014	TATCGTCGCGTGGTATTAACAGCTTGAGACATTAGTAGTAACAATTAACATCGCTGTG	209955
QY	481	TACATGAGGAGATCACTAAGTAAAGCCACAAACAATTTACTAATGCATTAACAACGCT	540
Db	209954	TTTAGCCTTGGTGCATGAATGAAGCCCTGTTGAAATTCGCAACAGCGCATCGAGCGGCG	209895
QY	541	AAAGAAGACATTTAGACATTTGTAATCATATGATACAGCAGTCGATACACATCGATGAA	600
Db	209894	AAAGAAGAGCATCACGACATATGTCTCATCGATACACAGCTGACGCTTCACATCGATGAA	209835
QY	601	GCATTGATGAACGAATTAAGAAGATTAAGAAATTCCTTAACCCAAACAATTAATGTTA	660
Db	209834	GAGCTGATGGATGAATCTCAACAAGTAAGGAAGTGCAGCAGCATGAATCTTGCTC	209775
QY	661	GTTCGCGATTCAATGACGGGTCAAGATGCTCAATGCTGCAGAACCTTTTGAGGATCAA	720
Db	209774	GTTCGTTGATGGAATGACAGGCGAAGATCGGTTATGTGCGGAAGCTTCATGAGCAG	209715
QY	721	CTTGATGTCACAGGTGTTTACCTTAACATAATAGATGATGATACAGTGGTGGTGCAGCT	780
Db	209714	CTCGATGTTTCCGCGCTGTTTTGACGAAGCTGGATGGGATACTCGAGCGGAGCTGCC	209655
QY	781	TTATCTATTCTGTCGTCACAAACAAATTAATTTCTTGATGATGATGATGAAGTTA	840
Db	209654	CTCTCGGTTAAGCGGTTAGAACACGCGCATCAATTTCCGCGGATGGTGAAGAATTT	209595
QY	841	GATGCTTTAGAGCTATTCCATCTCTGAAGTATGTCATCAGCTATTATTAGTATGGGTGAT	900
Db	209594	GACCAAGCTCGAACCGTTCCACCCAGAGCGAATGCTTCGCGGATCTTGGGATGGGGAT	209535
QY	901	GTGTTAAGTTTAATGTAAGAAAGCGCAACAAGATGTGATCAAGAAAGCAAAAGATTTA	960
Db	209534	GTGCTGACACTCATTGAAAAAGCAACATCCACTGTCGATGAAGAAAGCGCGTGAATT	209475
QY	961	GAGAAAGATGCGTGAGTATCGTTTACTTATAGATGATGATTTTGAACAACATGATCAG	1020
Db	209474	GAGAAAGAAATGCGTACATGACATGACATGCTGATGATTTCTTGAGGACATGACCAA	209415
QY	1021	GTGAAAAATCTAGACACACTGATGATATATGAAAAATGATCCAGGTATGAATAAATG	1080
Db	209414	GTGCGGACATGGGACCACTCGATGAATGCTGCGAATGATGCCAGGATGAACAAAGCA	209355
QY	1081	AAAGSGCTAGATAGCTTAATATGATGTAAGAAACAATTAATCATATTAAACCGATTATC	1140
Db	209354	AAAGGCTTGAAGAACTTGCAGTGGAGGAGAGCAAAATTAAGCGGCTTGAGCGGATGTC	209295
QY	1141	CAGTCAATGACCGCGCTGAAGAAACAATCCAGACACATTAATGATCATCGTAAAAAG	1200
Db	209294	CGTTCATGACGAAGCAAGAGAAACAAGATCTTAGCATCAATTCGAGTCTGCTCGA	209235
QY	1201	CGTATTGCTAAAGGCTGCTGCTTCATTAACAAGATCAATCGTTTGTATGAACAATTT	1260
Db	209234	CGGATCGGGAAGGAAGCGGAGCGGATTCAGATGTCACACCGCTGTTAAAGCGATT	209175
QY	1261	AACGATATGAAGAAATGATGAACAATTCAC	1292
Db	209174	GAAGATGAAGAAAGATGATGAACAATGAC	209143
RESULT	13		
LOCUS	AP003191	296750 bp	DNA linear BCT 10-JUL-2002
DEFINITION	Clostridium perfringens str. 13 DNA, complete genome, section 7/10.		
ACCESSION	AP003191	BA000016	
VERSION	AP003191.2	GI:18146729	
KEYWORDS			
SOURCE	Clostridium perfringens str. 13 (strain:13) DNA.		
ORGANISM	Clostridium perfringens str. 13		
	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
REFERENCE	1		

AUTHORS	Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A., Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
TITLE	Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE	21664373
PUBMED	11792842
REFERENCE	2 (bases 1 to 296750)
AUTHORS	Shimizu,T.
TITLE	Direct Submission
JOURNAL	Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical Sciences, University of Tsuba, Department of Microbiology; 1-1-1 Tennohdai, Tsukuba, Ibaraki 305-8575, Japan (E-mail:ts Shimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354, Fax:81-298-53-3354)
COMMENT	On Jan 14, 2002 this sequence version replaced gi:18145205.
FEATURES	Location/Qualifiers
Source	1..296750
	/organism="Clostridium perfringens str. 13"
	/strain="13"
	/db_xref="taxon:195102"
gene	/note="anaerobic pathogen for gas gangrene"
	complement(87..773)
	/gene="biob"
CDS	complement(87..773)
	/gene="biob"
	/note="228 aa, similar to sp:B10D_METJA DETHIOBIOTIN SYNTHETASE (EC 6.3.3.3) (DETHIOBIOTIN SYNTHASE) (DTB SYNTHETASE) (DTBS) from Methanococcus jannaschii (248 aa); 33% identity in 221 aa overlap
	CPE1543"
	/codon_start=1
	/transl_table=11
	/product="dethiobiotin synthetase"
	/protein_id="BAB81249.1"
	/db_xref="GI:18145206"
	/translation="MEKGVVLIIGTNDIGKTFISGLILKLRGGNAGYKAVLSGA IEKNGLLPDCSEVKKISGLKESYENMVSVILENPYSPHLASEVEVSISEKIKKD IYSDKXIDFILCGSGGIVCPISFSEKMLMEDIIRKFNQIILVNSGIGTINITYL LTVSVRLNGLKVGKIILNFKNSDIIHRDNKKIKELTGVNNISTVPIKEDIEKIDL NELNEVLVGI"
gene	complement(776..1735)
	/gene="biob"
CDS	complement(776..1735)
	/gene="biob"
	/note="319 aa, similar to pir:E70384 biotin synthetase (EC 2.8.1.6) from Aquifex aeolicus (332 aa); 43.4% identity in 311 aa overlap
	CPE1544"
	/codon_start=1
	/transl_table=11
	/product="biotin synthetase"
	/protein_id="BAB81250.1"
	/db_xref="GI:18145207"
	/translation="WDFILKMDKSLKNIKLTREGRLFNSNLEELKEANNIRKEI HDGDLGSLINGSGRCGDCAFCAQKYIKNTISEYPLLDYKIKKVAKEDEGY HRFISVTSRGVLGVEPERVITYYINLKNELKINLCASHGINKSLIKKAGVKRY HHNLETSRYVDKICKTHSYEPKVTIKNAEGLVCSGIIIGLGTILDRIDLIAT LRELEIKSIPINVLISAIGTKLQNLPIANESEILRTIAYPFINPEAKIRLAGGYLL KNFGNAFAGANATITGNLLTTCGNKINDDKRLIENIGMRIF"
gene	complement(1754..2308)
	/gene="biob"
CDS	complement(1754..2308)
	/gene="biob"
	/note="184 aa, similar to sp:BI0Y_BACSH BIOY PROTEIN from Bacillus sphaericus (215 aa); 43.4% identity in 173 aa overlap. Putative N-terminal signal sequence and 4 putative transmembrane regions were found by PSORT.
	CPE1545"
	/codon_start=1
	/transl_table=11
	/product="biotin synthase"
	/protein_id="BAB81251.1"

```
/db_xref="GI:18145208"
/translat="MKMKTRDLTFLIPFAALTAAGAFIKIPITPIVPFFLOVFCAQLGA
ILLAGKAGLAQIIVYVGLGLGIPVFTGGSPSVVQPTFGVLGIFGAVIIGKITE
NKKTNINKNIFLAGILGLIYVIFGVSYMYVYVIFLGDAMSLLSANVIGALCLPSD
LULFIISLVGSKVCPRLKRFDI1"
complement(2694..3785)
/gene="CPE1546"
complement(2694..3785)
/gene="CPE1546"
/note="363 aa, similar to sp:YPUA_BACSU HYPOPHETICAL 31.3
KDA PROTEIN IN LISA-PPIB INTERGENIC REGION (ORF19) from
Bacillus subtilis (290 aa); 31% identity in 271 aa
overlap. Putative N-terminal signal sequence was found by
PSORT"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB81252.1"
/db_xref="GI:18145209"
/translat="MRVLQGAIVSTFMFGMATPVFAQSEKTEAVVTLGANLTKSERLQ
MLDAFGVKAENEVIIDVTNODIREQLGIDTSKPIPASSQSISSVYVVKDGGINVT
NNLFEVTVGSANALLTSGVNADYKADAPKVTGAALAGILKGFDAAGEELSLPK
KEAREEISLTNNLSNAKTKDGTGKDEAAVVNDIKTDYIKDKPKNDERIGKIYNN
VTYNNILLTGGQEQETIKPMSKINDLDYNGAKESLAQMKQQLIKDTGKQLEE
SGLEKALNKKVLDIKDFLVNMFSSASEKVKDGIYDENGNIIVIKTNGNSDESKN
BESTQDPQTQSDNNQNGENEQGNQTK"
complement(4141..7578)
/gene="spovK"
complement(4141..7578)
/gene="spovK"
/note="1145 aa, similar to gp:AB015670_2 Bacillus sp.
genes for CDase, Cerase, MBP and 15 ORFs, partial and
complete cds from Bacillus sp (549 aa); 45% identity in
522 aa overlap. Also similar to prf:2516401X stage V
sporulation protein K from Bacillus cereus
CPE1547"
/codon_start=1
/transl_table=11
/product="stage V sporulation protein K"
/protein_id="BAB81253.1"
/db_xref="GI:18145210"
/translat="MLQKNIFVERKDVNKLINSGFFIEKKLNEEFIGQTKFNKEL
CDYFKIEEDSKGLIIVEERDIFKNSVLKIFFEELNYIKFKNSKIDEIDLAAYKF
NKNFATLFDLYSKIANDSQCLVFRNTDKASEDILNVLGISNCSIMLKNYVYKKN
FLVAEEIDEDIVNKLICNNKEFIFLYNKDGELETIDNEFLINKDKTLVKALSEK
DKNAYKKLLIKINKVKELFDIILIGVDENPDVEKSGVCTYIEDNFRKGSLSLD
CYLSKICKPLINIIRKEAIEANSRLIYVEDHIEICKSNDEVDLSKSNPTLEEYR
YKLSIIGMKELKEFDKIENNFVKQIREKGLKTKQSLNIFAGNAGTKNNAK
ITPFIILALGMIKENIFKEVSRADFIGEGISNTARRTNEIIESALGGVLFIDEAISC
TDENDKAGREIVDALIKGIEDNRNITVILAGYEKDMETFLSNQGLKSRFPNVIKFE
DYNSEMYEAVNIAKSGYRIAKAKAGLIELFKNQVKGSDIGNARFVRNTVENA
IMDAKKYLVNKRADILDEKDNFNKVSADFLEEKIEIIGLEVEKFLRSYKLI
VAOEKRSVGVNTKIBONLNFAPNPGTGKTSIARVAEMLNSMLGLLVQVLETDR
SSFYSDIPGETSKTEKFEKALGVFLIDEAYTLANDSIGREAITLLKLIEDYSE
VIVILAGTEEMENFDVNIIGLRSRFLPTKFDYFNELLEMALKEVAKFLKSN
GUSELKNFVEIENADAGSGMRVNVYEQILIRNQSIRIAESDLSVEMNLNKM
IKAMNAIKYDNDFELKLNLIIGNEELKFLRGYQIKMKIREKKKLGFPVDLYNM
NMVPTGSGCKTKVLNTIYSEMYLGIKAKNTVQIDKYEPMANISGMKVEDILNK
HVGLIYIDKWNTPFGBERYNEIVSDLIKFTDNNRNRIYVILGGIRGMKDLITNQG
LNYRFPVDEGYNERYELDIASSTLYVKGFTLDEBEATLENALGDIQKRMGELL
XNGLMIQFDSLVRSQIRVCEQDFDINTINRIISDILKSTFEIFLKNKTQNGSF"
complement(7634..9622)
/gene="fdh"
complement(7634..9622)
/gene="fdh"
/note="662 aa, similar to pir:S66017 formate dehydrogenase
homolog YyaE from Bacillus subtilis (667 aa); 32.1%
identity in 667 aa overlap
CPE1548"
/codon_start=1
/transl_table=11
/product="probable formate dehydrogenase"
/protein_id="BAB81254.1"
```

```
/db_xref="GI:18145211"
/translat="NVLISHGCTLDCDFCKCFNVYKEGSEILKIEGDKHEPHTKGLIC
KKGVAHLNRLNKHORYTPLLKNNGWEEISFEDALEIMKELEYTKYKSKSILYV
SOYSGSVLKGIEDIFFNPIYGVSKATGGFCWAGRAQKYDFGVSNSLEDMINSK
NIFUWGNPANTTITHMAILNARKKNSRIIVDINTOSAKLGDHVKIKPGVDGAL
AMAKIILISKGLQDKDFINKYVLGFQFYKDHLENFDLYLSDCEGIEIEDIKLY
PKNSSTIYLGVMQRYKNGGNTIRADALGALTQIGVKGGVYANKYANLILRSLOD
PYKEGSEVGNREFYVSNINETTEPKKYSLSVEDSNAPIKIWIYANSNMNLPNLNR
LNNSIDKPEKVCYCFOMEDTASKDLFIPCTNLTSESDMVFSSMTNPLYLYNEKLE
PREKLDIYFFRELAERMANLKGYPSSKKDYSKVIIEPLMSYNKDITFLYLNKNEFT
YVDDAVENKKFKTPSGFELASKRAKRCESLTPILSPRIKENCFLRLNHSKSL
SSQYIDVDKCAKVLNENIRFKSLTCEKVKLKSGTGEITAITACISLDNGVDYVALM
YVGVWKKHGNPNFLTESGIDMGGQITYNETFIEINI"
complement(11177..12238)
/gene="prs"
complement(11177..12238)
/note="364 aa, similar to pir:A71345 probable
phosphoribosyl pyrophosphate synthetase (prs) from
Treponema pallidum (421 aa); 32.3% identity in 325 aa
overlap
CPE1549"
/codon_start=1
/transl_table=11
/product="probable phosphoribosyl pyrophosphate
synthetase"
/protein_id="BAB81255.1"
/db_xref="GI:18145212"
/translat="NNEQNHGIIALESCTELGNAIDKLIQEKENCEESFLIKTDPI
RFSNGEKYAIPOSVNGRDIIYILCDIGNSYCTYKMGFENHKGDPDEHQDKRTVSAI
RGRARITVIMPLLYESROHRKRGESLDICALQLERLGVDEVTITFDVHDHPVQNA
LKFARFNFYPTDYIYKSLIHNEKDELDKELIVISPTGDMAIRATYISVLGVDVG
LFYKRRDHSIRVNGKPIVKHKEHGRVDGKDLIVDDMIASGESVLIDIAKELGRNA
RVNVATYFSFTEGLEKFNKFPNDGIKSVISTNLTYYIPELNAAPWFKAVDLSDFI
ARIINRLNVDKSVASVMDATNLIORLIDEX"
complement(11177..12238)
/gene="CPE1550"
complement(11177..12238)
/gene="CPE1550"
/note="353 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81256.1"
Query Match 38.2%; Score 522.8; DB 1; Length 296750;
Best Local Similarity 64.3%; Pred. No. 7.1e-79;
Matches 836; Conservative 0; Mismatches 432; Indels 12; Gaps 3;
QY 1 ATGSCATTTGAAGCGTTATCAGACGCGCTGCAAGCGACGATGCACAAAATCCGCTGAAG 60
DB 185046 ATGGCTTTTGATGATTAGCGCTCAAAATTACAGACACGCTTAAAAAATAAGGCTAAA 184987
QY 61 GGTAACCTTACGACGTATATAAGATATGATCGCTGAGTAAGATTAGCGTTATT 120
DB 184986 GGTAAAGCTAACGAAAAATATAAAAGCCATGAGAGAGTAAGCTTGGCGTTT 184927
QY 121 GAGCGTCAGCTAAACTTTAAAGTGGTAAAGAAATTTATTAACACATGATCAGACGCGCA 180
DB 184926 GAGCGAGNTGTACTTCAAGTGGTAAAAAGCTTTATATCAATGTTAAGATAATGCT 184867
QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACACGAGGCAACAAGTATTAAAAAGTCAA 240
DB 184866 GTTGGTCAGGAAGTATTAATAAGTTTAACTCCAGGTCACACAGCTTATTAATAAGTAA 184807
QY 241 GATGAATTAACCAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCATAATAACA 300
DB 184806 GAGCAATTAACAACCTCTTATGGGAGAGACTGAGAGTAATTAATAACTCAGACAATGA 184747
QY 301 CCYACTGTTGTTATGATGTTGTTTACAAAGTGGTGGTAAAAACAACACTGCAGGCTAAA 360
DB 184746 CCACACGTTTATGTTAGTAGTGTCTTCAAGGTGCGAGGTGCAAAACACTATGGCTGAAA 184687
```

QY	361	TTAGCATTTATGTCGGTAAATAACACAAAAACCTATGTTAGTTGCAGCAGATATT	420
Db	184686	CTTGCAATTGCATTAAAGAAAGAAA---AACAAAAAGCCATTACTGTGGCTTGTGATATA	184630
QY	421	TATCGTCCAGCAGCATAAATTAACAAACAGTAGGGAACAAATTTGATATTCCTGTA	480
Db	184629	TATAGACCTGCAGCATTAAGCAATTAACAGTTGTAGGAACAAATAGATATACAGTA	184570
QY	481	TACAGTGAAGGAGATCAAGTAAGACCAACAAATTTGTAACTAATGATTAATAACATGCT	540
Db	184569	TTCTCAATGGGAGTAAGGTTAAGGCTGTAGATATTGCCAAGCTGCAATAGAGCATGCA	184510
QY	541	AAAGAGAAACATTTAGACTTTGTAATCATTAACAGCAGGCTCGATTAACATCGATGAA	600
Db	184509	AAAGACATAGGAATAATGCTTGTATATAGTATACCGCTGGTAGACTTCATATAGATGAA	184450
QY	601	GCATGATGAAGCAATTAAGAAAGTAAAGAAATTCGTAACCAACAAAGAAATATGTTA	660
Db	184449	GATTTAATCAAGAGCTTAAGAGATGTTAAGAGATTTCAATCCTAGTGAATATATTATTA	184390
QY	661	GTTCGTGATTCATTAACGCGGTCAAGATGCTGTCATGTTGCGAATCTTTTGACGATCAA	720
Db	184389	GTATGTTGATGCTATGACAGGCTCAAGATGCTGTTAATSTAGCTGAGACATTTAATACTCA	184330
QY	721	CTTGATGTCACAGGCTTTACCTTAACCTAAGTAAATTAGATGGTGATACACGFGTGGTCAGCT	780
Db	184329	TTAGATCTTTCGGGAATATATTAAACAACTTAGATGGTGATACAGAGGTTGGTGTCT	184270
QY	781	TTATCTATTCGTTCCGGTGACACAAACCAATTAATTTGTTGTTGATGAGTGAAGTA	840
Db	184269	CTTTCATAAGACATTAACGTTAAACCAATTAATTTGTTGTTGTTGTTGTTGTTGTTGTTG	184210
QY	841	GATGGTTAGACGTTATTCATCCTTGACGTTATGCGCATCAGTATTTAGTATGGGTGAT	900
Db	184209	AGTGATATTGAAGTATTCATCCAGATAGATGGCTTCAAGATATATTAGGAATGGGAGAT	184150
QY	901	GTGTAAAGTTTAAATGAAGGCGCAACAGAGTGGATCAGAAAGCAAGCAAGATTTA	960
Db	184149	GTCCCTTCTTAAATAGAAAGCTCAGCAAGCTATTGACCAGATGAGCAAGTAAGTTA	184090
QY	961	GAGAAAGATCGGTGAGTCATCGTTTACGTTTAGTTAGATGATTTTATGACAACTTGATCAG	1020
Db	184089	AGTGAAAAAATGTTAAATCAAGAAATTTAACTTTGATGACTACTTATCAGCTATGGATCAA	184030
QY	1021	GTGAAATCTAGGACCTCGATGATATTATGAAAATGATTCAGAGTATGATATAAATG	1080
Db	184029	ATGAAAGCTTGGACCTATAAATAAATGATAGATGATTCAGAGGTTTACACAAAG	183970
QY	1081	AAAGGGCTAG--ATAAGCTTAATATGAGTGAAGCAAAATGATCATATTAAAGCGATT	1137
Db	183969	GAACCTTGAAGGCATTGATTTTCTCAAGGAGAAAAACAAATGGCTACAGTTAAAGCTATC	183910
QY	1138	ATCCAGTCAATGACGCGGCTGAAGAAACAATCC-----AGACCAATTAAGTATCA	1191
Db	183909	ATACAATCAATGACAGCTTAAAGAAAGAAAAACAACTTCCTTTAGTAATAGGAATGTTCA	183850
QY	1192	CGTAAAAAGCGTATGCTAAAGGCTCGTCTCATTAAGAGAGTCAATCGTTTGATG	1251
Db	183849	AGAAAGNAGAGATAGCTTAAGGTTCTGGTACACAGTACAGAAAGTAATTAAGTCTTA	183790
QY	1252	AAACAATTTAACGATATGAAGAAAATGATGAACAAATCA	1291
Db	183789	AAAGGCTATGAATGATGAAGAAAGCAAAATGAAGCAATCA	183750

RESULT 14
AE010040/c
LOCUS
DEFINITION
Streptococcus pyogenes strain MGAS8232, section 88 of 173 of the complete genome.
ACCESSION
AE010040 AE009949
VERSION
AE010040.1 GI:19748305
KEYWORDS

SOURCE	Streptococcus pyogenes MGAS8232.
ORGANISM	Streptococcus pyogenes MGAS8232 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE	1 (bases 1 to 10173)
AUTHORS	Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M. genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks
TITLE	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
JOURNAL	21927593
MEDLINE	11917108
PUBMED	
REFERENCE	2 (bases 1 to 10173)
AUTHORS	Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M. Direct Submission Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St., Hamilton, MT 59840, USA
TITLE	Location/Qualifiers
JOURNAL	1..10173
FEATURES	/organism="Streptococcus pyogenes MGAS8232" /strain="MGAS8232" /db_xref="taxon:186103" complement(380..1450) /gene="spvM18_1147" complement(380..1450) /gene="spvM18_1147" /note="best blast match gb AAK34058.1 (AE006560) putative integrase/recombinase [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative integrase/recombinase" /protein_id="AAL97765.1" /db_xref="GI:19748306" /translation="MRRELLLEKIETKAIMPWVLYVYQSKLAIPYSFTTLYLVEYKRFNRLMDADLTQAPKTADTDLSTLEHLTKKDLRAFVLYLRERPSLNTYSTREGISOTTIRLSLSLYLYLPEEVENDQCEPFYRNVKKVSTKKKXETLASRAENIKOKLFLGDTALFDYDKEIEQKLSNRKASSFFKNRKNRDLAIIALLASGVLRSEAVNDLKDYNMMIIIEVRGGRDSVNVAGFAGYLESLAVRORRYKAERQDIAFLFTBYRGVNRMDASSIEKMWKTSDFKTRVTPHKLRHTLATRLYDATKSOVLVSHQLHSSQTQVDLYTHIVNDEQKNALDNL" 1913..2611 /gene="spvM18_1150" 1913..2611 /gene="spvM18_1150" /note="best blast match gb AAK34059.1 (AE006560) putative repressor protein [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative repressor protein" /protein_id="AAL97766.1" /db_xref="GI:19748307" /translation="NPSGHQKTKARLSKIGTQSELGRLLHVNKMTISNWEKGNKIPNEKHLNALLFNFTSDCFDNRYLLTPYNOLTIENKSKVIGYSERLLNHQIDKSKDLIDKPSQLYAFVYESLSAGDGYSGFDGNFDVYDQLESDYFASWVFGDSMEPTPLNGEVLIKONSFYDGIYAVEMDQGTIRKKVFRDEGLRLVLSLNKKYISDKFAPSEPRITIGKTIANFRPLEI" complement(2746..4308) /gene="ffh" /note="spvM18_1151" complement(2746..4308) /gene="ffh" /note="best blast match gb AAK34060.1 (AE006560) putative signal recognition particle -inhibited division protein [Streptococcus pyogenes M1 GAS]"

```
/codon_start=1
/transl_table=11
/product="putative signal recognition particle -inhibited
division protein"
/protein_id="AAL97767.1"
/db_xref="GI:19748308"
/translation="MAFESLQRLQDVFKIRGKKLSQSDVQVETKEIRIALLEADV
ALPVVFTIKRVRAIGHEIIDIDPTKOILKIVNEETILGSEVAFIDKSPKPT
IIMVGLQAGATTFAGLANKLIIKEEDARPLMAADIRPAADQLKTDGQGINVPT
FMDGTHSDAVIVRGLQROARENENDVLDTAGRLQIDDKIMBELRDVALAQPNEI
LIVVDSMIGQEAANYAVENHOLSITGVLTIKDGTGGGALSVELTCKPKTKFTGT
GERITDIETPHDPMSSRLIENGDLHLIPEKASQEDDEKKLSLEAKRMRENTDFDND
LEQDQVQNGPMDOLLKXIPGMAGNPALANIKVDENQIARKRAIVSSMPABRENDP
KNPAGNPDMSSLEGMGCGMPDLSLGGMDMSQLFGGFGKGIQGFAMQAMKRPQ
ANKLKAKKKRR"
complement(4323..4664)
/gene="yImx"
/note="spym18_1152"
complement(4323..4664)
/gene="yImx"
/note="best blastp match gb|AAK34061.1| (AE006560)
putative DNA-binding protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="AAL97768.1"
/db_xref="GI:19748309"
/translation="WNIMEIKTRNALFEFVAALLTDKQNNYIELYYADDVSLAEI
EAGFVSQRVDNIRKTEKILETYEMKHMVSDYVVRSEIFDDMIAHYPDEYLOEK
ISLTSIDNRE"
complement(4753..5451)
/gene="spym18_1153"
complement(4753..5451)
/gene="spym18_1153"
/note="best blastp match gb|AAK34062.1| (AE006560)
putative transcription regulator, GntR family
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcription regulator"
/protein_id="AAL97769.1"
/db_xref="GI:19748310"
/translation="MLPAYIKIHDAIKKIDIDGIWPIGSRLPSEHLEHFTVSRMTL
ROATLLVEGILERRIGSTVYASHRVQEKMRGTTFTFELIRSQGPSSKLLSYOK
QLASDTYKELNDLTDLIRMERIRYADSVPLVYELASIEPEKIKTVKRADITEFF
HSLANGYIGKSKQTIYAKLASERVASYLEVAKGAILALTQVSIFYTDGKPEYVRS
QIGDRFEFYLENN"
complement(5534..6967)
/gene="spym18_1154"
complement(5534..6967)
/gene="spym18_1154"
/note="best blastp match gb|AAK34063.1| (AE006560)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL97770.1"
/db_xref="GI:19748311"
/translation="MLRKKRLGLFRVELIFGLLLSLVSYLAWNSPATLHNILA
TVGIVERSKQOPRHIGQATQVQKSGFYEQWGTINKQVEDIAENTRVSYHYVTFP
IGKVTYSLPEHLKRPDKPRFKGDIVKLSLTKEPHIKVYQGLATIKQVKRYDYP
LGQYDGINLKDLDJDSBQDFVKPIYIRFNKNSPQNNRLIRKAFAYAKOHPNS
VISPGQGHISLSPKDYFELSDPAILIGHOTFEILHGKMLWFGPTGPKAEQVRS
NLVLTGVHKANDLKKGDHFMIMADHGTDEHVIYDNKFWHKRNSHIFDLGSLNSLF
EKNQFIGYAPELVQQLLSKAQGHDFSEVIOFDVAHHFAWDGGLSLNIAPNYAF
NORTHLCHNITYSQNOLPYIDPTGCLRAYSGSIGHSKGVIRVLNNVFTSSIVTK
AKLTSWFMEPIHFPNPSPIVAVAGNIIN"
7134..8696
/gene="guaa"
/note="spym18_1155"
7134..8696
```

```
/gene="guaa"
/note="best blastp match gb|AAK34064.1| (AE006560)
putative GMP synthase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative GMP synthase"
/protein_id="AAL97771.1"
/db_xref="GI:19748312"
/translation="MTEISILDVQKTIIVLDYGSOVNOLIARRIRERGVSSELSKSHI
TAQELREINPIGIVLSGSPNSVADNAPGIDPEIFELPILGICYGQMLITHLKGSK
VPYAGAGNREGGQSTLHLRETSKLFQEQQLVMSHGDAVTEIFEGHVLGSDND
CPVAATENKMLYIGIOFHPVRHSVYVDNLIKNFATISCGARGDSNDFIDMEIAK
IRTVGDKRVLLGLSGDSSVGVLLQKAIQGLQTCIFVDHGLLRDKDEGDQVGMGLG
GKFLNLIRVDASKRFLDLLADVEDPEKSKIIIGNEFVYVFDKASLKGVDLFAOGT
LYTDITESTETAOIKSHNNVGLPEDMOFELIBPLNLFKDEVRALGIALGMPBEI
VWQPGPGCLAIRVWGALTEELFTVRESDAILREIAKAGLDROVQWQVFTVTVYR
SVGVMGDGRITYYTAIRATISIDGMTADPAQLPMDVLKAKISTRIVNEVDHVRIVD
ITSKPPATVEWE"
complement(8738..9961)
/gene="spym18_1156"
complement(8738..9961)
/gene="spym18_1156"
/note="best blastp match gb|AAK34065.1| (AE006560)
putative antimicrobial resistance factor [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative antimicrobial resistance factor"
/protein_id="AAL97772.1"
/db_xref="GI:19748313"
BASE COUNT 3112 a 1914 c 1748 g 3399 t
ORIGIN
Query Match 36.5%; Score 499.2; DB 1; Length 10173;
Best Local Similarity 61.3%; Pred. No. 1.5e-74;
Matches 804; Conservative 0; Mismatches 508; Indels 0; Gaps 0;
QY 1 ATGGCAATTGGAAGCTTATCAGAACGCTCGAAGCGAGATGCAAAAATGCGGTAG 60
Db 4308 ATGGCTTTTGAAGCTTAACCCACGCTTTGCAAGATGTTTAAACACATTCGCGGTA 4249
QY 61 GGTAACCTTACTGAAGCTGATAAAGATAATGATCGGTGAAGTAAGATTAGCGTTATT 120
Db 4248 AAGAAATTATCAGATCAGATGTTCAAGAAGTGACAAAAGAGATTCGCTAGCTTTTA 4189
QY 121 GAGCGTACGTAACCTTAAAGTGGTAAAGAATTATTATAAAGATGATCAGAACGCGA 180
Db 4188 GAAGCGATGTCGCTTTACCAGTGTGTTAAACCTTTTATTAAGCGCGCTACGTGAGC 4129
QY 181 TTAGGTCGATGTAATGCAATCATTAACACCGGCGCAACAGTTATTAATAAGTCAA 240
Db 4128 ATTGGCCATGAGATTATGATCTCTTGATCCAAACAAATTCCTTAAGATTGTAAT 4069
QY 241 GATGAATTAACGAAGTTGATGGTGAGGAAAATACATCGATTATATGTCAAATAACCA 300
Db 4068 GAAGAATTAACGAGTATTTAGCTCAGAACTCGAGAGATTGATAATCTCAAAAT 4009
QY 301 CCTACGTTGTTATGAGTGTGTTTACAGGTGCTGCTGAACACACAACTGCAGSTAA 360
Db 4008 CCAACTATTATCATGATGTCGCTTTGCAAGGGCTGCTGAACACAACTTTCAGSAAA 3949
QY 361 TTAGCATTTATGATCCGCTAAAATAACAAACAACTTACTGATTCGACAGATATT 420
Db 3948 CTCGCTAATAATAATAAAGAAGAGATCGCGACCATGATGATGTCGCGCATATT 3889
```


gene	/db_xref="GI:1850606"	
	/translation="MEYEKTRNNALFEYFAALLRDKOMNYIELYVADDSYLAETAE	
	EDVSRQAVDNKIRTEKLEDEYKMLHMYSDYVSESEFDALMKKYPNDPYLQNKISI	
CDS	LPTINDRD"	
	969..2519	
	/gene="ffh"	
gene	969..2519	
	/gene="ffh"	
	/notes="similar to 54 kDa subunit of signal recognition	
CDS	particle"	
	/codon_start=1	
	/transl_table=11	
gene	/product="frn"	
	/protein_id="AAB48050.1"	
	/db_xref="GI:1850607"	
CDS	/translation="MAPESLTERLQGVFNLRGKRLSEKDVQVYKEIRLALLADV	
	ALPVKFEIKRVRKRAVHEVIDLDPSSQIIKIIVNEELTAVLGSETAEIKSSKIPT	
	IIMVGLGAGTFFAGKLANKLVRENARPLMIADIRPAADQLKILGOQINVPV	
gene	FMGTGHEAVEIVSQLAQAKENRNDYVLIDTAGRLQIDEXLMTFELRDILKALANEI	
	LIVDSMIGQERANVAREFNOQLVGVILIKIDGTRGGAALSVROITGPKTKFTGT	
	GKTIIDTFHPDRMSSRIILMGDLJTLIEKASODYDQKSAELAEKRENSFDNDF	
CDS	IEGLDQVQMSMEDILKIPGMANKPALANVEYDEGEIARKRALVSSMTPERENPD	
	LTPSRRIASGSGTIVNVNKFIKDFNQARKMGVSGDMNKVHKQGINPNRPD	
	NNMDSALBGMGOGMPDMKSLGSANMDVSMFGGLGKGVGEFAMKSNKKAKRM	
gene	KKAKEKK"	
	2539..3066	
	/gene="satC"	
CDS	2539..3066	
	/gene="satC"	
	/note="unknown function"	
gene	/codon_start=1	
	/transl_table=11	
	/product="SatC"	
CDS	/protein_id="AAG28335.1"	
	/db_xref="GI:11055963"	
	/translation="MNLHNTYSIKRYNGLPNKIQIIIIALSLSIYLPLILLISN	
gene	GVNGDLHFHIDRMLGLSIWKSYPNFTFYKGVGVNYFYFYRYKGVGVVDNRL	
	VSSALSKGSVAIHPNGIHEIRVYVYAVARVAFAISIIIVIIIVYVLYHLEK	
	HRTTEKIRGLFIY"	
CDS	3163..3831	
	/gene="satD"	
	3163..3831	
gene	/gene="satD"	
	/note="unknown function"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="SatD"	
	/db_xref="GI:11055964"	
gene	/translation="MYIAIIGDLISSKATTNPKSOKLNLLNOLNKKYKELLKSA	
	FTTTGDFQALLVPNPQIFOLIDEIALGPKPYLRFVGSSSILITINPEQSIGSDG	
	PAYWHARAIDYIHDKNYGSNHLAYDLETDSQINAILACEFIKSKWYTYQYEV	
CDS	IDGLLQGIYEKFSHKXMAEKLDLSPSSFNKKLSSGLKIYLRNKKVATILLNLAIR	
	KERR"	
	3828..>4152	
gene	/gene="satE"	
	3828..>4152	
	/gene="satE"	
CDS	/note="unknown function"	
	/codon_start=1	
	/transl_table=11	
gene	/product="SatE"	
	/protein_id="AAG28337.1"	
	/db_xref="GI:11055965"	
CDS	/translation="MTKLVISQFLSGNFVILLILIAHFLADFQWQSKMADLKSNW	
	TYLIRHLIIVALLPILLSVPHSFILSLIFLSHLVDSGLKLLNSFYKDRSFIKTK	
	AVFLID"	
BASE COUNT		1412 a 664 c 781 g 1295 t
ORIGIN		
Query Match		36.3%; Score 497.2; DB 1; Length 4152;
Best Local Similarity		61.2%; Pred. No. 4e-74;
Matches 802; Conservative		0; Mismatches 508; Indels 0; Gaps 0;

Qy	1081	AAAGGCGTAGATAGCTTAATATGAGTGGAAAGCAAAATTGATCATATTAAAGCGATTATC	1140
Db	2049	CCAGCTTGTGCCAATGTGCGAAGTTGATGAAGTGGAAATTTCTAGAAAACGGCGCCATCGTA	2108
Qy	1141	CAGTCANTGACGCGCTGGAAGAACAAATCCAGACACATTTGAATGATCATCGTFAAAAAG	1200
Db	2109	TGATCAATGACACTGAAGAAGCTGAGAAATCCTGACCTCTAACTCCGAGGCGGACAGCT	2168
Qy	1201	CGTATTGTTAAAGGCTCGTCTGCTTCATCAAGAAAGTCAATCGTTGATGAAGAACAATTT	1260
Db	2169	CGATTGTCTTGGTTCTGCTAATACTTTTGTAAATGTCATAAATTCATCAAGAAATTT	2228
Qy	1261	AACGATATGACAGAAAATGATGAAAACAAATTCACCTCGTCCGCGGTAAAGGTAA	1310
Db	2229	AATCAGGCTTAAGAAAGATGATGCAAGGTGTCTGCTGCTGATGATGAATAA	2278

Search completed: February 24, 2003, 23:49:03
Job time : 5411 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 20:04:35 ; Search time 260 Seconds

(without alignments)
11848.976 Million cell updates/sec

Title: US-09-943-108A-1

Perfect score: 1368
Sequence: 1 atgcaattgaagcttattc.....gtatgaattaccgttttaa 1368

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1368	100.0	1368	20	AXX27221 S. aureus ffh codi
2	1356.8	99.2	1368	23	AAS5331 Staphylococcus aur
3	1344	98.0	1368	23	AAS54393 Staphylococcus aur
4	1341	98.0	1365	23	AAS51783 Staphylococcus aur
5	790.4	57.8	792	20	AXX27222 S. aureus ffh codi
6	743.2	54.3	1014	24	ABN92495 Staphylococcus epi
c 7	703.4	51.4	721	23	AAS49372 Staphylococcus aur
8	700.8	51.2	319630	24	ABQ67194 Listeria innocua c
c 9	697.6	51.0	2256	24	ABQ70511 Listeria monocytog

c	10	612.4	44.8	644	18	AAV75400 Staphylococcus aur
	11	568.6	41.6	1416	23	AAS51308 Enterococcus faeca
	12	568.6	41.6	1434	23	AAS53127 Enterococcus faeca
	13	565	41.3	6729	20	AAI12974 Staphylococcus aur
c	14	516.2	37.7	521	23	AAS49408 Staphylococcus aur
c	15	516.2	37.7	521	23	AAS49436 Staphylococcus aur
	16	505.6	37.0	1563	24	ABN67087 Streptococcus poly
	17	496	36.3	1395	23	AAS55790 Streptococcus pneu
	18	496	36.3	1569	20	AAI19484 Streptococcus pneu
	19	496	36.3	1569	24	ABK48379 DNA encoding strep
	20	496	36.3	1572	23	AAS55519 Streptococcus pneu
	21	496	36.3	1605	22	ABH90837 2CFE 28 coding seq
c	22	496	36.3	7577	19	AAV52179 Streptococcus pneu
	23	492.8	36.0	1572	22	ABH90724 CFE 28 coding sequ
	24	491.2	35.9	1563	24	ABN67088 Streptococcus poly
	25	486.4	35.6	1136	24	ABK74556 Bacillus lichenifo
c	26	477.6	34.9	2026	19	AAZ96317 S. pneumoniae deri
c	27	477.6	34.9	2026	19	AAV42992 Streptococcus pneu
c	28	463.2	33.9	468	23	AAS50493 Staphylococcus aur
c	29	443.4	32.4	1389	23	AAS53231 Haemophilus influe
c	30	443.4	32.4	1830121	17	AAT42063 Haemophilus influe
	31	433.8	31.7	533	18	AAV75697 Staphylococcus aur
	32	418.2	30.6	640681	24	ABA92787 Buchnera sp. genom
	33	404	29.5	92407	22	AAF28549 Genomic fragment #
	34	370.6	27.1	1362	22	AAH84550 E. coli growth and
	35	370.6	27.1	1362	23	AAS52529 E. coli DNA for ce
	36	360	26.3	1506	23	AAS56354 Salmonella typhi D
	37	341.4	25.0	8367	21	AAH81483 N. meningitidis pa
c	38	341.4	25.0	349980	21	AAF21544 Neisseria meningit
c	39	341.4	25.0	349980	21	AAF21607 Neisseria meningit
c	40	328.4	24.0	330	23	AAS49670 Staphylococcus aur
	41	280	20.5	1374	23	AAS54213 Pseudomonas aerugi
c	42	279.4	20.4	910715	20	AAZ02448 Borrelia burgdorfe
	43	268	19.6	4804	20	AAV65141 Ehrlichia sp. E74
c	44	263.6	19.3	292	23	AAS50792 Staphylococcus aur
	45	260.8	19.1	1230025	20	AAV91990 Nucleotide sequenc

ALIGNMENTS

RESULT 1
AXX27221
ID AAX27221 standard; DNA; 1368 BP.
XX AAX27221;
AC AAX27221;
XX
DT 28-MAY-1999 (first entry)
XX
DE S. aureus ffh coding sequence.
XX
FFh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;
KW immunological response; gene therapy; infection; otitis media;
KW conjunctivitis; toxic shock syndrome; septic arthritis; ss.
XX
OS Staphylococcus aureus.
XX
PN EP902087-A2.
XX
PD 17-MAR-1999.
XX
PF 24-AUG-1998; 98EP-0306741.
XX
PR 10-SEP-1997; 97US-0927216.
XX
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Wallis NG;
XX
DR WPI; 1999-169238/15.
DR P-PSDB; AAY00910.
XX
PT New Staphylococcus aureus Signal Recognition Particle (SRP) with

Dd 1201 CGTATTCCTAAAGGTTCTGGTCGTTTCATACAGAAAGTCAATCGTTTGATGAACAATTT 1260

QY 1261 AACGATATGAAGAAAATGATGAACAATTCACCTGGTGGCGGTAAAGGTAAAGGTAA 1320
|||||

Dd 1261 AACGATATGAAGAAAATGATGAACAATTCACCTGGTGGCGGTAAAGGTAAAGGTAA 1320
|||||

QY 1321 CGCAATCAAAATGCAAAATATGTTAAAGGATGATGAATTTACCGGTTTAA 1368
|||||

Dd 1321 CGCAATCAAAATGATGTTAAAGGATGATGAATTTACCGGTTTAA 1368
|||||

RESULT 4
AAS51783
ID AAS51783 standard; DNA; 1365 BP.
XX AAS51783:
XX
Dt 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus DNA for cellular proliferation protein #200.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206948P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsep KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX
DR WPI; 2001-611495/70.
XX
PT P-PSDB; AAU33924.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Claim 27; Seq ID No 4365; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 1365 BP; 518 A; 190 C; 285 G; 372 T; 0 other;
Query Match 98.0%; Score 1341; DB 23; Length 1365;
Best Local Similarity 98.9%; Pred. No. 9.8e-290;
Matches 1350; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGCGATTTGAAGGCTTATCAGAACGCTCGAAGCGAGATGCAAAATTCGCGTGAAG 60
|||||

Dd 1 ATGCGATTTGAAGGCTTATCAGAACGCTCGAAGCGAGATGCAAAATTCGCGTGAAG 60
|||||

QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATTT 120
|||||

Dd 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTGAAGTAAGATTAGCGTTATTT 120
|||||

QY 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180
|||||

Dd 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180
|||||

QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACGAGCGCAACAGTATTAAATAGTTCAA 240
|||||

Dd 181 TTAGGTTCCGATGTAATGCAATCATTAACACGAGCGCAACAGTATTAAATAGTTCAA 240
|||||

QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATATCGATTAAATGCTCAATFAAACCA 300
|||||

Dd 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATATCGATTAAATGCTCAATFAAACCA 300
|||||

QY 301 CCTACTGTTGTATGATGGTGGTTTACAAGGTGCTGGTAAACAAACACACTGCAAGTAAA 360
|||||

Dd 301 CCTACTGTTGTATGATGGTGGTTTACAAGGTGCTGGTAAACAAACACACTGCAAGTAAA 360
|||||

QY 361 TTAGCATTATTGATGCGTAAATAATACAAACAAAACCTATGTTAGTTGCAGCAGATATT 420
|||||

Dd 361 TTAGCATTATTGATGCGTAAATAATACAAACAAAACCTATGTTAGTTGCAGCAGATATT 420
|||||

QY 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAAATTGATTTCTGTGA 480
|||||

Dd 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAAATTGATTTCTGTGA 480
|||||

QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACCTAATGCAATTAACACATGCT 540
|||||

Dd 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACCTAATGCAATTAACACATGCT 540
|||||

QY 541 AAAGAAGAACATTTGACATTTGTAATCATGATACAGAGGTGCGATTACACATCGATGAA 600
|||||

Dd 541 AAAGAAGAACATTTGACATTTGTAATCATGATACAGAGGTGCGATTACACATCGATGAA 600
|||||

QY 601 GCATTGATGAACGAATTAAGAAGATTAAGAAGAAATTTGCTAAACCAACAAATTTATGTTA 660
|||||

Dd 601 GCATTGATGAATGAATTAAGAAGATTAAGAAGAAATTTGCTAAACCAACAAATTTATGTTA 660
|||||

QY 661 GTTGTGATTCANTGACGGGTCAAGATGCTGCAATGTTGCGAGAACTTTTGACGATCAA 720
|||||

Dd 661 GTTGTGATTCANTGACGGGTCAAGATGCTGCAATGTTGCGAGAACTTTTGACGATCAA 720
|||||

QY 721 CTTCATGTCACAGCTTTACCTTAACTAAATAGATGCTACACAGTGGTGGTGCAGCT 780
|||||

Dd 721 CTTCATGTCACAGCTTTACCTTAACTAAATAGATGCTACACAGTGGTGGTGCAGCT 780
|||||

QY 781 TTATCTATTTCGTTCCGTGACAAAACCAATTAATTTGTTGGTATGAGTGAAGTAA 840
|||||

Dd 781 TTATCTATTTCGTTCCGTGACAAAACCAATTAATTTGTTGGTATGAGTGAAGTAA 840
|||||

QY 841 GATGGTTTACAGCTATTCACCTCGTGAAGTATGTCATCGATGTTTATGAGTATGGGTGAT 900
|||||

Dd 841 GATGGTTTACAGCTATTCACCTCGTGAAGTATGTCATCGATGTTTATGAGTATGGGTGAT 900
|||||

QY 901 GTGTAAAGTTTAAATTCAAAAGCGCACACAGATGTGATCAAGAAAAGCAAAAGATTTA 960
|||||

Dd 901 GTGTAAAGTTTAAATTCAAAAGCGCACACAGATGTGATCAAGAAAAGCAAAAGATTTA 960
|||||

QY 961 GAGAAAAGATGCGTGATCGTTTACTTTAGATGATTTTTTGAACAACATTTGATCAG 1020
|||||

Dd 961 GAGAAAAGATGCGTGATCGTTTACTTTAGATGATTTTTTGAACAACATTTGATCAG 1020
|||||

QY 1021 GTGAAAATCTAGACACACTGGATGATATTATGAAATGATTCAGGTATGATTAATG 1080
 Db 1021 GTGAAAATTTAGGACCACTGGAGATATTAATAATGATTCAGGTATGATTAATG 1080
 QY 1081 AAAGGCTAGATAGCTTAATATGATGAGTAAAGCAAAATGATCATATTAAGCGATTATC 1140
 Db 1081 AAAGGCTAGATAGCTTAATATGATGAGTAAAGCAAAATGATCATATTAAGCGATTATC 1140
 QY 1141 CAGTCAATGACCGCGCTGAAAGAAACAATCCAGACACATGATGATCAGGTAAAG 1200
 Db 1141 CAGTCAATGACCGCGCTGAAAGAAACAATCCAGACACATGATGATCAGGTAAAG 1200
 QY 1201 CGTATTGCTAAAGGCTGCTGCTGCTCATTAACAAGATCAATGCTGATGAAACAATTT 1260
 Db 1201 CGTATTGCTAAAGGCTGCTGCTGCTCATTAACAAGATCAATGCTGATGAAACAATTT 1260
 QY 1261 AACGATATGAAGAAAATGATGAACAATTCATCTGCTGCGGTAAAGGTAAAGGTAAA 1320
 Db 1261 AACGATATGAAGAAAATGATGAACAATTCATCTGCTGCGGTAAAGGTAAAGGTAAA 1320
 QY 1321 CGCAATCAATGCCAAATATGTTAAAGGTATGATGAATTTACCGTTT 1365
 Db 1321 CGCAATCAATGCCAAATATGTTAAAGGTATGATGAATTTACCGTTT 1365

RESULT 5
 AAX27222
 ID AAX27222 standard; DNA; 792 BP.
 XX
 AC AAX27222;
 AC
 XX
 DT 28-MAY-1999 (first entry)
 XX
 DE S. aureus ffh coding sequence.
 XX
 KW Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;
 KW immunological response; gene therapy; infection; otitis media;
 KW conjunctivitis; toxic shock syndrome; septic arthritis; ss.
 XX
 OS Staphylococcus aureus.
 XX
 PN BP02087-A2.
 XX
 PD 17-MAR-1999.
 XX
 PF 24-AUG-1998; 98EP-0306741.
 XX
 PR 10-SEP-1997; 97US-0927216.
 XX
 PA (SMIK) SMITHKLINE BEECHAM.
 XX
 PI Wallis NG;
 XX
 DR WPI; 1999-169238/15.
 XX
 DR P-PSDB; AAY00911.
 XX
 PT New Staphylococcus aureus Signal Recognition Particle (SRP) with
 PT protein (ffh) and RNA (ffs) components - the SRP gene and protein
 PT useful as diagnostic reagents and for prevention and treatment of
 PT Staphylococci infections which cause otitis media, septic arthritis
 PT and toxic shock syndrome
 XX
 PS Claim 1; Page 30-31; 35pp; English.
 XX
 CC This sequence encodes the Staphylococcus aureus signal recognition
 CC particle (SRP) ffh component. Ffh polynucleotides are useful for
 CC diagnosing a disease related to expression of ffh polypeptides by
 CC analysing for the presence/amount of ffh protein in a sample due to
 CC infection of a micro-organism with the gene, or determining the nucleic
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful
 CC for treatment of an individual in need (polypeptide) of, or needing to
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and

CC polynucleotides are useful for identifying agonists and antagonists by
 CC binding and observing the affect of ffh polypeptide activity, which are
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments
 CC are also useful for inducing an immunological response (T cell/antibody)
 CC to protect against disease, by direct administration (vaccine), or via a
 CC vector (gene therapy). Anti-ffh antibodies are useful as antagonists, and
 CC for protecting against disease. Diseases diagnosed, prevented and treated
 CC include those caused by infection, especially bacterial infection,
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and
 CC are useful for bathing wounds and implants prior to surgical
 CC implantation.
 XX

SQ Sequence 792 BP; 296 A; 116 C; 164 G; 216 T; 0 other;

Query Match 57.88; Score 790.4; DB 20; Length 792;
 Best Local Similarity 99.9%; Pred. No. 6e-167;
 Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCAATTGAAGGCTTATCAGAAACGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60

Db 1 ATGGCAATTGAAGGCTTATCAGAAACGCTGCAAGCGAGATGCAAAAATGCGTGTAAAG 60

QY 61 GGTAAACTTACTGAAGCTGATATAAAGATAATGATGCGTGAAGTATGAGTATTT 120

Db 61 GGTAAACTTACTGAAGCTGATATAAAGATAATGATGCGTGAAGTATGAGTATTT 120

QY 121 GAGGCTGAGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACCGCA 180

Db 121 GAGGCTGAGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACCGCA 180

QY 181 TTAGGTTCCGATGTAATGCAATCATTAAACACAGGCGCAACAGTTATTAATAATGATCAA 240

Db 181 TTAGGTTCCGATGTAATGCAATCATTAAACACAGGCGCAACAGTTATTAATAATGATCAA 240

QY 241 GATGAATTAACGAAGTGTGCGGTGGAGAAATACATCGATTAATATGTCATAATAACCA 300

Db 241 GATGAATTAACGAAGTGTGCGGTGGAGAAATACATCGATTAATATGTCATAATAACCA 300

QY 301 CCTACTGTTTGTATGATGTTGTTTACAGGTGCTGTTTAAACCAACAACCTGAGTAAA 360

Db 301 CCTACTGTTTGTATGATGTTGTTTACAGGTGCTGTTTAAACCAACAACCTGAGTAAA 360

QY 361 TTAGCAATTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGACGAGATAT 420

Db 361 TTAGCAATTATGATGCGTAAATAATACAAACAAACCTATGTTAGTTGACGAGATAT 420

QY 421 TATCGTCCAGCAGCGATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480

Db 421 TATCGTCCAGCAGCGATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480

QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATGTAACTTAATGCAATTAACCAATGCT 540

Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATGTAACTTAATGCAATTAACCAATGCT 540

QY 541 AAAGAAGACATTTAGACTTTGTAATCATGATACAGAGGTGCGATTAACATCGATGAA 600

Db 541 AAAGAAGACATTTAGACTTTGTAATCATGATACAGAGGTGCGATTAACATCGATGAA 600

QY 601 GCATTGATGAACGATTTAAAGAGATTAAGAAATTTGCTAAACCAACGAAATATGTTA 660

Db 601 GCATTGATGAACGATTTAAAGAGATTAAGAAATTTGCTAAACCAACGAAATATGTTA 660

QY 661 GTTGTGATTCATGACGGGTGACAGATGCTGTCATGTTGCAAGATCTTTTTCAGCATCAA 720

Db 661 GTTGTGATTCATGACGGGTGACAGATGCTGTCATGTTGCAAGATCTTTTTCAGCATCAA 720

QY 721 CTTGATGTCACAGGTTTACCTTAACATAATAGATGCGTATACACGTTGCGTGCAGCT 780

Db 721 CTTGATGTCACAGGTTTACCTTAACATAATAGATGCGTATACACGTTGCGTGCAGCT 780

QY 781 TTATCATTCGTT 792

Db 781 TTATCATTCGTT 792

Db 781 TTAUCTATTGGT 792

RESULT 6
ABN92495
ID ABN92495 standard; DNA; 1014 BP.
AC ABN92495;
XX
XX 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.
KW
XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
XX
XX P-FSDB; ABP39950.
XX
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 1958; 267pp; English.
XX
XX ABN90538 to ABN9374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;

Query Match 54.3%; Score 743.2; DB 24; Length 1014;
Best Local Similarity 84.1%; Pred. No. 2.2e-156;
Matches 838; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 1 ATGGCATTGAAGCTATACAGAGCCTGCAAGCGATGCAAAAATGCGGTGAAG 60
|||||
Db 16 ATGGCATTGAAGGATTATCCGATGCTTACAGCCACGATGCAAAAATGCGGTAAA 75
|||||
QY 61 GGTAACTACTGAAGCTATACAGAGTAAAGATATGATGCTGAAGTATAGATGCGTTATT 120
|||||
Db 76 GGAAGATTAACAGAGCAGAGATTAACAGATGATGCTGAAGTATAGATGCGTTATT 135
|||||
QY 121 GAGCGTACGTAACCTTTAAAGTGTAAAGAAATTTATTAACAGTATACAGAGCGCA 180
|||||
Db 136 GAAGCGGATGTTAACTTCAAGATGTTTAAAGAAATTTTGAAGATGTTTCAAGAGCG 195
|||||
QY 181 TTAGTTCGATGTAATGCAATCATACACAGCGGCAACAGTATTAATAATAGTTCAA 240
|||||
Db 196 CTAGGTTCTGATGTCATATTTAAACCTGGCCACAGGTTATTAATAATGTCACAA 255
|||||

QY 241 GATGAATTACGAGAGTGTGATGGTGGAGAAATACATCGATTAAATATGCAAAATACCA 300
|||||
Db 256 GAAGAAGCTTACTAGTTTAAATGGTGGAGAAATACTTCCATTAAAGATGGCAAAATACCA 315
|||||
QY 301 CCTACTGTTTATGATGTTGTTTACAAGGTGCTGGTAAACCAACAACTGCGAGTAAA 360
|||||
Db 316 CCAACTGTTGTCATGATGCTGGCTTACAGGTCAGAGTAAACCCAGCAGCAGGTAAG 375
|||||
QY 361 TTACATATTGATGCGTAAATAATACACAAAACCTATGTTAGTTGCGACGAGATTT 420
|||||
Db 376 TTGCATTATTATGCGTAAATAATATACAAAACCTTTACTTTGCGCAGGAGATTT 435
|||||
QY 421 TATGCTCCACGCGATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480
|||||
Db 436 TATGCTCCAGCTCTATTGATCAATTCAACACAGTAGTAAACAAATGATATCCCTGTG 495
|||||
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAACATAATGCATTTAAACATGCT 540
|||||
Db 496 TATAGTGAAGGTGATCAAGTATCACCAACAAATTTGTTGAAAATGCTTTAAACATGCT 555
|||||
QY 541 AAAGAAGACATTTAGACTTTGTTAATCATTTGATACAGCAGGTGATACATCGATGAA 600
|||||
Db 556 AAAGAAGACATCTAGATTTCGTTATCATTTGATACAGCTGCTGTTTACATTTGATGAA 615
|||||
QY 601 GCATTGATGAACGAATTAAGAAGTAAAGAAATTCGTAACCAACGAAATTTATGTTA 660
|||||
Db 616 GCATTTATGATGAGCTTCAAGAAGTTAAGAATCTTAACCCAGCAGAAATTTATGCTT 675
|||||
QY 661 GTTGTGATTCAATGACGGTCAAGATGCTGTAATGTCAGAAATCTTTGACGATCAA 720
|||||
Db 676 GTTGTGATGCAATGACAGGTCAAGATGCTGTAATGTCAGCAATCATTTGATGAA 735
|||||
QY 721 CTGATGTCACAGGTGTTTACCTTAATTAATAGATGCTGATACAGGTGCTGCTGATG 780
|||||
Db 736 TTAGATTGTTTCAAGTGTAACTGTTAATTAATAGATGCTGATACAGCGGTGCTGAC 795
|||||
QY 781 TTATCTATTGTTGCTGACACAAAACCAATTAATTTGTTGGTATGATGAGTAAAGTTA 840
|||||
Db 796 CTTCAATCGTTCCTGACCAAAACCTTAATTAATTTGATGATGATGAGTAAAGTTG 855
|||||
QY 841 GATGTTTATAGAGTATTCCTGATGATGATGATGATGATGATGATGATGATGATGAT 900
|||||
Db 856 GATGTTTATAGATTTATTCCTGATGATGATGATGATGATGATGATGATGATGATGAT 915
|||||
QY 901 GTGTTAGTTTATTTGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGATGAT 960
|||||
Db 916 GTTTTAGCTCTTATTTGAAAGCGCAACAGATGATGATGATGATGATGATGATGATGAT 975
|||||
QY 961 GAGAAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
|||||
Db 976 GAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
|||||

RESULT 7
AAS49372/c
ID AAS49372 standard; DNA; 721 BP.
XX
XX AAS49372;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Staphylococcus aureus cellular proliferation inhibitory sequence #596.
DE
XX
XX Antisense; ss; prokaryotic cellular proliferation;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX W0200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
PF

XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 28-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS
XX Claim 1; Seq ID No 1949; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence is an antisense
CC oligonucleotide of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 721 BP; 191 A; 149 C; 104 G; 277 T; 0 other;

Query Match 51.4%; Score 703.4; DB 23; Length 721;
Best Local Similarity 99.2%; Pred. No. 1.5e-147;
Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGGCTTATCAAGACCGCTGCAAGCGACGATGCAAAAAATCGGTGTAAG 60
DB 713 ATGGCATTTGAAGGCTTATCAAGACCGCTTGCAGCGACGATGCAAAAAATCGGTGTAAG 654

QY 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCTGAGTAAGTAAGTATGCTTATTT 120
DB 653 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCTGAGTAAGTAAGTATGCTTATTT 120

QY 121 GAGGCTGACGTAACATTTTAAAGTGGTAAAAAGATTTTATTAACACATGATCAGAACGCGCA 180
DB 593 GAGGCTGACGTAACATTTTAAAGTGGTAAAAAGATTTTATTAACACATGATCAGAACGCGCA 534

QY 181 TTAGGTCGATGTAATGCAATCATTAACACAGGCGCAACAGTATTAAAAAGTTCAA 240
DB 533 TTAGGTCGATGTAATGCAATCATTAACACAGGCGCAACAGTATTAAAAAGTTCAA 474

QY 241 GATCAATTAACGAGTTGATGGTGGAGAAATACATCGATTAAATGCTCAATTAACCA 300
DB 473 GATCAATTAACGAGTTGATGGTGGAGAAATACATCGATTAAATGCTCAATTAACCA 414

QY 301 CCTACTCTTGTATGATGTTGGTTTACAAGGTGCTGTTAAAAACAACACTGCAGGTAAA 360
DB 413 CCTACTCTTGTATGATGTTGGTTTACAAGGTGCTGTTAAAAACAACACTGCAGGTAAA 354

QY 361 TTAGCATTAATGATGCGTAAAAAATACAAACAAAAACCTATGTTAGTTGACGAGATATT 420

DB 353 TTAGCATTAATGATGCGTAAAAAATACAAACAAAAACCTATGTTAGTTGACGAGATATT 294
QY 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATTGATATTCCTGTA 480
DB 293 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATTGATATTCCTGTA 234
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAACATAATGCAATTAACACATGCT 540
DB 233 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAACATAATGCAATTAACACATGCT 174
QY 541 AAAGAGACACATTAAGACTTTGTAATCATGATACAGCAGGTGATACACATGATGTA 600
DB 173 AAAGAGACACATTAAGACTTTGTAATCATGATACAGCAGGTGATACACATGATGTA 114
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGTAACCAACAAACGAATTAATGTTA 660
DB 113 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGTAACCAACAAACGAATTAATGTTA 54
QY 661 GTTGTGATTCATCAATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGA 713
DB 53 GTTGTGATTCATCAATGACGGGTCAAGATGCTGCAATGTTGCGAATCTTTTGA 1

RESULT 8
ABQ67194
ID ABQ67194 standard; DNA; 319630 BP.
XX AC ABQ67194;
XX 29-AUG-2002 (first entry)
DI Listeria innocua contig DNA sequence #7.
DE Listeria innocua
XX KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
OS Listeria innocua.
XX PN WO200228891-A2.
XX PD 11-APR-2002.
XX PF 04-OCT-2001; 2001WO-FR03061.
XX PR 04-OCT-2000; 2000FR-0012697.
XX PA (INSP) INST PASTEUR.
XX PI (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX PS Claim 5; SEQ ID 7; 180pp; French.
XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 319630 BP; 105207 A; 55428 C; 66726 G; 92263 T; 6 other;

Query Match 51.2%; Score 700.8; DB 24; Length 319630;
 Best Local Similarity 70.7%; Pred. No. 2.5e+146;
 Matches 933; Conservative 0; Mismatches 387; Indels 0; Gaps 0;

QY 1 ATGGCAATTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAATCGTGSTTAAG 60
 DB 187924 ATGGCAATTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAATCGTGSTTAAG 187983

QY 61 GGTAACTTACTGAAGCTGATTAAGATTAATGATGGGTGAAGTGAAGTGAAGTATTT 120
 DB 187984 GGTAACTTACTGAAGCTGATTAAGATTAATGATGGGTGAAGTGAAGTGAAGTATTT 188043

QY 121 GAGCTGACGTAAACCTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 180
 DB 188044 GAGCTGACGTAAACCTTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 188103

QY 181 TTAGGTTCCGATGATTAATCAATCATTAACACGAGGCAACAGTATTAAATAGTTCAA 240
 DB 188104 TTAGGTTCCGATGATTAATCAATCATTAACACGAGGCAACAGTATTAAATAGTTCAA 188163

QY 241 GATGAATTAACGAAGTTGGTGGGAGAAATACATCGATTAATATGTCAAATAAACCA 300
 DB 188164 GATGAATTAACGAAGTTGGTGGGAGAAATACATCGATTAATATGTCAAATAAACCA 188223

QY 301 CCTACTGTTGTTATGATGTTTGTGTTTCAAGGTCGCTGTTAAACACACACTGCGAGTAAA 360
 DB 188224 CCTACTGTTGTTATGATGTTTGTGTTTCAAGGTCGCTGTTAAACACACACTGCGAGTAAA 188283

QY 361 TTAGCATATTGATGCGTAAAAATACAAACAAACCTATGTTAGTTCGACGAGATATT 420
 DB 188284 TTAGCATATTGATGCGTAAAAATACAAACAAACCTATGTTAGTTCGACGAGATATT 188343

QY 421 TATGTCGACGACGATTAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA 480
 DB 188344 TATGTCGACGACGATTAATCAATTAACACAGTAGGGAACAAATGATATTCCTGTA 188403

QY 481 TACAGTGAAGAGATCAAGTAAAGCCACACARATTTTACTAATGCAATTAACACATGCT 540
 DB 188404 TACAGTGAAGAGATCAAGTAAAGCCACACARATTTTACTAATGCAATTAACACATGCT 188463

QY 541 AAAGAAGACATTTAGACATTTTGTATCATTTGATACAGAGTTCGATTCACATCGATGAA 600
 DB 188464 AAAGAAGACATTTAGACATTTTGTATCATTTGATACAGAGTTCGATTCACATCGATGAA 188523

QY 601 GCATTGATGAACGATTAAGAGATTAAGAAATTCCTTAACACCAAGAAATGATTTA 660
 DB 188524 GCATTGATGAACGATTAAGAGATTAAGAAATTCCTTAACACCAAGAAATGATTTA 188583

QY 661 GTTGTGATTCATGACGCGTCAAGATGCTGTCATGTTGACAGATCTTTTGAAGATCAA 720
 DB 188584 GTTGTGATTCATGACGCGTCAAGATGCTGTCATGTTGACAGATCTTTTGAAGATCAA 188643

QY 721 CTTGATGTCAGGTTTACCTTAACTAAATAGATGGTGATACAGTGGTGGTGCAGCT 780
 DB 188644 CTTGATGTCAGGTTTACCTTAACTAAATAGATGGTGATACAGTGGTGGTGCAGCT 188703

QY 781 TTACTATTCTGTCGGTGACACAAACCAATTAATTTTGGTATGATGAGTAAAGTTA 840
 DB 188704 TTACTATTCTGTCGGTGACACAAACCAATTAATTTTGGTATGATGAGTAAAGTTA 188763

QY 841 GATGTTTGAAGTATTCCTCTGAAAGTATGCAATTCACGATTTTAAAGTATGGTAT 900
 DB 188764 GATGTTTGAAGTATTCCTCTGAAAGTATGCAATTCACGATTTTAAAGTATGGTAT 188823

QY 901 GTGTTAAGTTAATGAAGAGCGCAACAGATGATGATCAAGAAAGCAAAAGATTTA 960
 DB 188824 GTGTTAAGTTAATGAAGAGCGCAACAGATGATGATCAAGAAAGCAAAAGATTTA 188883

QY 961 GAGAAAGAGCGTGATGATCGTTTACTTTAGATGATTTTGAACAACATTCATCAG 1020
 DB 188884 GAGAAAGAGCGTGATGATCGTTTACTTTAGATGATTTTGAACAACATTCATCAG 188943

QY 1021 GTGAAAACTAGACCACTGATGATATTAATGAAATGATTCAGGATGATGAATAATG 1080
 DB 188944 GTGAAAACTAGACCACTGATGATATTAATGAAATGATTCAGGATGATGAATAATG 189003

QY 1081 AAAGGGCTAGATGAAGCTTAATGATGAGTGAAGCAAAATGATCAATTAATAAGCGATTATC 1140
 DB 189004 AAAGGGCTAGATGAAGCTTAATGATGAGTGAAGCAAAATGATCAATTAATAAGCGATTATC 189063

QY 1141 CAGTCATGACGCGGCTGAAAGAAACAAATCCAGACACATGTAATGATCACTGTAAGAG 1200
 DB 189064 CAGTCATGACGCGGCTGAAAGAAACAAATCCAGACACATGTAATGATCACTGTAAGAG 189123

QY 1201 CGTATTCCTAAAGGCTGCTGCTGTTCAATCAAGAGTCAATGCTTTGATGAACAATTT 1260
 DB 189124 CGTATTCCTAAAGGCTGCTGCTGTTCAATCAAGAGTCAATGCTTTGATGAACAATTT 189183

QY 1261 AACGATATGAAGAAATGATGAACAAATTCCTGCTGGCGGTAAGGTAAGGTAAGGTAAG 1320
 DB 189184 AACGATATGAAGAAATGATGAACAAATTCCTGCTGGCGGTAAGGTAAGGTAAGGTAAG 189243

RESULT 9
 ABQ70511/C
 ID ABQ70511 standard; DNA; 2256 Bp.
 XX
 AC ABQ70511;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria monocytogenes 4b contig DNA sequence #453.
 XX
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 infection; ds.
 XX
 OS Listeria monocytogenes 4b.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 treatment and prevention of infection, also related polypeptides,
 antibodies and modulators -
 XX
 PS Claim 14; SEQ ID 3324; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 (ABQ7188-ABQ71212) from Listeria sp. The sequences are useful as probes
 and primers for identification and/or detection of Listeria (e.g. as
 contaminants in foods, or mutational analysis) and for analysis of
 gene expression. Proteins encoded by the nucleic acid sequences can be
 used to screen for compounds that modulate gene expression, replication
 and pathogenicity of Listeria (potential therapeutic agents), also for
 treating infections by Listeria, and are useful as immunogens in
 anti-Listeria vaccines.
 CC
 CC Note: The sequence data for this patent did not form part
 of the printed specification, but was obtained in electronic format
 directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 2256 Bp; 533 A; 471 C; 420 G; 832 T; 0 other;
 SQ

Query Match 51.0%; Score 697.6; DB 24; Length 2256;

Best Local Similarity 70.5%; Pred. No. 3.9e-146;
Matches 931; Conservative 0; Mismatches 389;

QY	1	ATGGCATTTGAAGCCTTATCAGAAGCCTGCAACGGACGATGCATAAAATCGTGTTAAG	60
Dd	1666	ATFGCGATTGAAGGACTAGCTGGAGACTCCCAAGAACATGAACAAAATTCGGGGCAA	1607
QY	61	GGTAACTTACTGAAGCTGATATAAGATTAATGATGGTGAAGTAGATTACGCTTAATT	120
Dd	1606	GGGAANGTAACGAAGCTGACCTTAAGAAATGATGGTGAAGTTCGCTTGCTCTACTT	1547
QY	121	GAGGCTGACGTAAACTTTAAAGTGTAAAAAATTTATTTAAACAGATATCAACGGCCA	180
Dd	1546	GAAGCCGATGTAACTTTAAAGTCGTAAACAATTTATTAARACAGTAGAGCAAGTCT	1487
QY	181	TTAGTTCGGAGTAAAGCAATNTACACCAGGCCAACAGTTATTAATATAGTTCAA	240
Dd	1486	GTCCGCGGGACGTTATGAAGAGCCTTAACCCGGGCCAACAAAGTTATCAAAATGCTTCAA	1427
QY	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCGATTAAATATGTCAAAATAACCA	300
Dd	1426	GAAGAACTTACAGCTTTAATGGCGGAGAGAAAGCAAAATCGGAACAGCGCACCGCCA	1367
QY	301	CCTACTGTGTATGATGTTGTTTACAAGTCTCTGTAAAAACAACACTATGTTAGTTCACGAGATTT	420
Dd	1306	CTTGCTAAATTTATTACGCAAAAATATTAACCGTAAACCATTTACTTCGCGACAGATATT	1247
QY	421	TATCTGTCAGCAGCATTAATCAATTAACAACAGTAGGGAACAAATATGATATTCCTGTA	480
Dd	1246	TACCGAGCTGCCGAATCAACACTAGAACACTTGGGAAGCAATTAGATATGCCGTA	1187
QY	481	TACGTGAAGGAGATCAAGTAAAGCCACACAAATTTTAACATTAATCAITTAACAACTGCT	540
Dd	1186	TTTTCTCTFAGGAGACCAAGTAAAGCCGGTAGAAATGCCAAACAACACTATTGCTAAAGCA	1127
QY	541	AAGAGAACAACTTATAGACTTTGTATCATCTGATACAGCGTCTGATACACATCGCTCA	600
Dd	1126	AAGAGAACAACTTATAGACTTTGTATCATCTGATACAGCGTCTGATACACATCGCTCA	1067
QY	601	GCATTTGATGAACGAATTAAGAAAGTAAAGAAATTCGTAAACCAACGAAATTAAGTTA	660
Dd	1066	ACGCTAATGGATGATTAACACAGTGAAGAAATTCACACGCCACCGAATCTGCTT	1007
QY	661	GTTGTCGATTCAAATACGCGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTTGAGCATCAA	720
Dd	1006	GTTGTCGATTCATGATCGGCAAGATGTGTCAACGTGGCGCAAGACTTCAACGAACAA	947
QY	721	CTTGATGTCACAGCTGTTACCTTTACTTAATTTAGATGGTATACAGTCTGCTGCTGCACT	780
Dd	946	TTAGAAATACAGGTGTCGTCTCAACAAATTTAGACGCTGATACAGTGGTGGTGCAGCG	887
QY	781	TTATCTATTTCGTTGGTGACACAAAACCAATTAATTTGTTGGTATGASTCAAAAGTTA	840
Dd	886	CTTTCGATTCGTTCCGCTACTGGCAACCAATTAATTTATCGSACTGTTGAAAAAATG	827
QY	841	GATGGTTTAGAGCTATTCCATCTCAACCTATGGCAACAGTATTTTAGTATGGGTGAT	900
Dd	826	GAAGCTTAGAAACTTTCCATCCAGATCTGATGGCTCCAGAATTCCTGGCATGGCGCAT	767
QY	901	GTCTTAGCTTTAATGAAAAGSCAACAGATGTGGATCAAGAAAAACCAAAAGATTTA	960
Dd	766	GTGCTTCCCTGATTGAAGAAGSCAAACCGATGTAGATACAGAAAAATGAAGACTATG	707
QY	961	GAGAAAAAGATGCGTGAGTCAATGTTTACTTTTAGATGATTTTTTAGAACAACTTCATCAG	1020
Dd	706	GAACAAAAATGAAGACACACATGACGCTAGATGACTTCTTAGNACAGTTACACAA	647
QY	1021	GTGAAAAATCTTAGGACCACTGGATGATATTATGAATATGATCCAGGTATGAATATAATG	1080

Db	646	GTAAACAATAATGGGACCACTAGATGAACACTACTTAAATGATGCCAGGCGCAACAACAATG	587
QY	1081	AAAGGGCTAGATAAGCTTAAATATGATGAAAGCAAATATGATCATATTAAAGCGATTATC	1140
Db	586	AAAGCCCTCGACAACATGACGTCGATGATAAACAACCTAGGTCATATTGAAGCAATTATC	527
QY	1141	CAGTCATGACCGCGCTGAAGAACAACATCCAGACACACATGAACTGATCATCGTAAACAAG	1200
Db	526	AAGTCANTGACAAAATGAAAGAACAAATCCGATATCATCAATGAAGTAGAAGAAA	467
QY	1201	CGTATTGCTAAAGGGTCGTGCTGCTCATTACAAGAAGTCAAATCGTTTGATGAACAATTT	1260
Db	466	CGAATGCACTGGTACCGGCGCTCGGTTCAAGAAATTAACCGTCTCCTAAACAATTT	407
QY	1261	AACGATATGAAGAAAATGATGAACAATTCACTGGTGGCGGTAAAGGTAAAAAAGGTAAA	1320
Db	406	GGTGAATATGAAAAAATGATGAAGCAAAATGACTGGTGGCGGCAAGGCAAAAAGGCAAAA	347

RESULT 10
AAV75400/c
ID AAV75400 standard; DNA: 644 BP.
XX
XX
XX AAV75400;
XX
XX
DT 16-MAR-1999 (first entry)
XX
XX
DE Staphylococcus aureus contig SEQ ID #1089.

Computer readable medium; vaccine; *S. aureus* infection; immunodetection
 xii cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 XX
 OS *Staphylococcus aureus*.

PN	EF786519-A2.	
XX		
PD	30-JUL-1997.	
XX		
PF	07-JAN-1997; 97EP-0100117.	
XX		
PR	05-JAN-1996; 96US-0009861.	
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;	
PI	Rosen CA;	
PI		
XX	WPI; 1997-374922/35.	
DR		
XX		
XX		
PT	Polynucleotide(s) and proteins derived from Staphylococcus aureus	
PT	stored on computer readable medium and used in the production of	
PT	anti-S.aureus vaccines	
XX		
PS	Claim 1; Page 1828; 3271pp; English.	

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the *S. aureus* DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against *S. aureus* infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating

301 CCGACAGTGAATTATGATGACACAGG

301 CCGACAGTGAATATGATGACAGGGTTACAAAGGGGCTGGTAAACAACCTTTACTGGTAAA 360

QY 361 TTACGATATTGATCGCTAAAAAATACACAAAAACCTATGTTAGTTGCGAGCATATTT 420
Db 361 TTACGAAAACATCTATGAAACTGAAAAGCTCGTCCGCTTTTATCGCTGGTGAAGTT 420
QY 421 TATCGTCCAGCAGCATTAATCAAAACAGTAGGGAACAATGATATTCCTGTA 480
Db 421 TATCGTCCAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 481 TACAGTGAAGGAGATCAAGTAAGCCACAAATGTAACATAATGATTAACACATGCT 540
Db 481 TTTGATATGGACAGATGCTAATCCAGTGGAAATGTTCTGTCAGGGTTAGCATTAGCA 540
QY 541 AAAGAGACATTTAGCTTTGTAATCATGATACACAGAGTCTGATACATCATGATGAA 600
Db 541 AAAGAAAGAAAATGANTATCTCTTAATGATACGCGCGCGTGTACACATTCAGCAA 600
QY 601 GCATGTGATGAACGAATTAAGAGAGTAAGAGAAATGCTTAACCCAAAGCAATTAAGTTA 660
Db 601 GCTTTAATGACGAATTAAGAAATTAAGAGATGGCTAATCCCAATGAAATTCGTGTA 660
QY 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCATGTTGCCAGAAATCTTTTGACATCAA 720
Db 661 GTTGTGATGCGATGACGGGCAAGATGCTGTCACGTTGCGAGATGTTTAAATGAACAG 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACATAATGATGATGATGATGATGATGATGATGAT 780
Db 721 CTTGGAATTAAGTGGGTGTTATTAACCAAAATGGACGGGATGATGATGATGATGATGAT 780
QY 781 TTATCTATTCGTTGCGTGACACAAAACCAATTAATTTGTTGATGATGATGATGATGAT 840
Db 781 CTGTCATTCGCGAGTACCGGGGCTCCGATTAATTAATTTGTCGGTTCTGGTGAAAAATTA 840
QY 841 GATGGTTTAGAGTATTCATCTCCGTAACGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 ACCGATTAGAAATTTCCATCCCGATGCTGATGTCGAGTCGATCTCTAGTATGSGGGAC 900
QY 901 GTGTTAAGTTTAAATGAAAAGGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960
Db 901 ATGTTGACGCTAATGAAAAGCGCAACAGATTAAGATGATGATGATGATGATGATGATGAT 960
QY 961 GAGAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTTAGACAACTTGATCAG 1020
Db 961 GCTCAAAAATGAAAGAAAACAGTTTGGACTTTAAGCATTTCAATGACCAATTCGATCAA 1020
QY 1021 GTGAAAATCTAGGACCATGGATGATATTAATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 GTTATGGCATGGACCGGATTAAGACTTATTAATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 AAAGGGTAGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 CCTGGTATGAAAATGTCAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 CAGTCAATGACCGGCTGAAAGAAAACATCCAGACACATGAAATGATGATGATGATGATGAT 1200
Db 1141 CTATCAATGACCGCTGACACCGTGAATATCCCTGATCTATTAATCTAGTCGCGCGTCG 1200
QY 1201 CGTATGCTAAAGGCTGCTGGTTCATTTACAAAGAGTCAATCGTTTATGATGATGATGATGAT 1260
Db 1201 AGAATTCGACCTGGTTACAGAAATAGTGTGGTGAAGTCAATCGTATGATGATGATGATGAT 1260
QY 1261 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1303
Db 1261 AAAGAAATCCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1303

RESULT 12

AAS53127

ID AAS53127 standard; DNA; 1434 BP.

XX AAS53127;

AC AAS53127;

XX 13-FEB-2002 (first entry)

XX

XX

XX

DE Enterococcus faecalis DNA for cellular proliferation protein #555.
XX Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX Enterococcus faecalis.
OS
XX WO200170955-A2.
PN 27-SEP-2001.
PD
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
DR P-FSD5; ANU35268.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX Claim 27; Seq ID No 6764; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1434 BP; 500 A; 230 C; 340 G; 364 T; 0 other;

Query Match 41.6%; Score 568.6; DB 23; Length 1434;
Best Local Similarity 64.8%; Pred. No. 2.1e-117;
Matches 844; Conservative 0; Mismatches 459; Indels 0; Gaps 0;

QY 1 ATGCGATTTGAAGGCTTATPACAGAGCGCTGCAAGCGACGATGCAAAAATCGGTGTAAG 60
Db 1 ATGCGCTTTTGAGAGTTTAAACAAACCGCTACAAACAGGCAATGAGTAAAAATCCGTCGTAAG 60
QY 61 GGTAACTTACTGAAGCTGATATTAAGATATGATGCTGAAGTATGATGATGATGATGATGAT 120
Db 61 GGAAGAGTTTCCGAGCGCGACGTAATAAGAAATGATGCGAGAAATCCGTTTGGCTTTATTA 120
QY 121 GAGGCTGACGTAACCTTTAAAGTGTGAAGAAATTTATTAACACAGATATCAAGAACGGCA 180
Db 121 GAAGCCGACGTTTAAATTTACAGAGTGTCAAGAGATTTCAAAAACCGCGTCAGAGAACGGCA 180
QY 181 TTAGGTTCCGATGTAATGCAATCATTAACACCGAGGGCAACAAAGTTATTAAAAATAGTTCAA 240

Db	181	GTAGGAGTCGAAGTATTAGAAAGCTTATCACACGCCACAAATATGTAATAATTTGTTGAT	240
Qy	241	GATGAATTAACGAAGTTGATGGGTGGAGAAATACATCGATTAAATTAATATGTCAAATAAACA	300
Db	241	GAAGATTAACGAACAAGTTAGTCTCAGAAACGGTTGACTGAATAAATCTCCAAATC	300
Qy	301	CTTACTGTGTGTTATGATGGTTGGTTTACAAGTGCTGTGTAAAAACAACAACACTCGAGTAAA	360
Db	301	CGACAGTGATTATGATGACAGGGTTTACAAAGGGCTGTGTAAAAACAACCTTTTACTGCTAAA	360
Qy	361	TTAGCATTAATTCATGCGCTTAAAAATACACAAAAAACCCTATGTTAGTTCACACAGATT	420
Db	361	TTAGCAAAACACTTAATGAATACTGAAACGCTGTGCGGCTTTTAATCGCTGGTGAOCTT	420
Qy	421	TATCGTCCACAGCAGTAATAATCAATTACAAACAGTAGGGAACAATAATGATATTTCTCTGTA	480
Db	421	TATCGTCCACAGCAGTATGATCAGTTGAGGTTTTAGGTCAACAATTAAGAATTCCTCGTT	480
Qy	481	TACGTGAAGGAGATCAAGTAAAGCCACACAATAATGTTAACTAATGCATTAATAACATGCT	540
Db	481	TTTGTATATGGGAACAGATGCTAATCCAGTGGAAATTTGTCGCAAGGTTAGCATAGCA	540
Qy	541	AAAGAAAGACATTTTAGACTTTTGTATTCATGATACACAGGTCGATACACATCGATGAA	600
Db	541	AAAGAAAAAATAATGATTATGCTTTTAATGATAGCGCGGCGGTTTACACATTCACGAA	600
Qy	601	GCATTGATGAACGAATTAAGAAGTAAAGAAATTTGCTAAACCAACCAAAATTAATGTTA	660
Db	601	GCTTAAATGGACGAATTAACAACAATAAGAGTTGGCTAATCCCAATGAAATTTCTGTTA	660
Qy	661	GTTGTGATTCAAATGACGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTTGACGATCAA	720
Db	661	GTTGTTGATCGGATCACGGGCAAGATGCTGTCAACGTTGCAGATGTTTAAATGAACAG	720
Qy	721	CTTGATGTCACAGGTGTTACTTTAACTAAATTAGATGGTGATACAGTGGTGCTGACGCT	780
Db	721	CTTGGAATTTACTGGGGTGTATTACAAAATGGAACGGCGATACCTCGTGGGGGGCTGCG	780
Qy	781	TTATCTATTTCGTTGCTGACACAAAAACCAATTAATTTGTTGSTATGAGTCAAAAGTTA	840
Db	781	CTGTCAATTCGGGCAAGTAACGGCGCTCCGATTAATTTTGTGCGTTCTCGTGAATAATTT	840
Qy	841	GATGGTTTAGAGCTATTCCATCCTGAAAGTATGSGCATCAGATTTTAGGTATGGGTGAT	900
Db	841	ACGGATTTAGAAATTTTCCATCCGATCGTATGTCGAGTCTATCCTAGGTAATGGGAC	900
Qy	901	GTGTTAAGTTTAATGTAAAAAGCGCAACAAGTGTGGATCAAGAAAAACAAGATTTA	960
Db	901	ATGTTGACGCTAATGTAAAAAGCGCAACAAGTATACGATGAGAAAAAACAACAAGACTT	960
Qy	961	GAGAAAAAGATCGGTGAGTCATCGTTTACTTTTAGATGATTTTTTAGAACAACTTCATCAG	1020
Db	961	GCTCAAAAATGAAGAACAACATTTTGACTTTACGATTCATGAGCAATGGATGCA	1020
Qy	1021	GTTAAAAATCTAGGACCACTGGATGATATTATGAAATGATTTCCAGGTATGATAAATG	1080
Db	1021	GTTATGGGATGGGACCGATTTAGAGACTTTATTAATAATGATCCCTGGAATGAGTAACATG	1080
Qy	1081	AAAGGCTAGATAGCTTAATATAGTGAAAAAGCAAAATGATCATATTTAAACGCAATTATC	1140
Db	1081	CTTGGTATTGAAAAATGTCAAAGTCGATCCAAAAGATGTGGCAGGAAACGGCGCATGGTC	1140
Qy	1141	CAGTCAATGACCGCGGTGAAGAAACAATCCACAGACATTTGAATGATCACTGTAANAAG	1200
Db	1141	CTATCAATGAOCCCTGCAAGACGTGAATACTGTATTTAAATCCTTAGTCCGCTGCG	1200
Qy	1201	CGTATTGCTAAAGGCTCGCTTCATTACAAGAAAGTCAATCGTTTGATGAACAATTT	1260
Db	1201	AGATTTGACGTGTTTACGAAATAGTGTGTTGAAGTCAATCGATGATTTAAACAATTT	1260
Qy	1261	AACGATATTGAAGAAATGATGAACAATTTCACTGGTGGCGGTA	1303

Dd	1261	AAGAATCCAAAAAATGATGCACAAATGTCCAAAGGGGATA	1303
	RESULT 13		
	AAx12974		
ID	AAx12974	standard; DNA; 6729 BP.	
XX	AC		
XX	AAx12974;		
XX	AC		
XX	AAx12974;		
DT	19-MAR-1999	(first entry)	
XX			
DE	Enterococcus faecalis	genome contig SEQ ID NO:37.	
XX			
XX	Enterococcus faecalis;	contig; detection; Enterococcal infection;	
KW	vaccine; attenuation;	computer readable medium; ds.	
XX			
OS	Enterococcus faecalis.		
XX			
XX	WO9850555-A2.		
PN			
PD	12-NOV-1998.		
XX			
PF	04-MAY-1998;	58WO-US08985.	
XX			
PR	14-NOV-1997;	97US-0066009.	
PR	06-MAY-1997;	97US-0044031.	
PR	16-MAY-1997;	97US-0046655.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Barash SC, Dillon PJ, Kunsch CA;		
XX			
DR	WPI; 1999-045171/04.		
XX			
PT	New isolated Enterococcus faecalis	polynucleotides and polypeptides	
PT	- used to develop products for the	detection of Enterococcus and for	
PT	use in vaccines for prevention or	attenuation of Enterococcus	
PT	infection.		
XX			
PS	Claim 1; Page 406-409; 2084pp;	English.	
XX			
CC	A computer readable medium has been	developed which has recorded on it	
CC	982 nucleotide sequences isolated	from the Enterococcus faecalis genome	
CC	AX12938 to AX13919 represent these	nucleotide sequences which are	
CC	primary nucleotide sequences, also	known as contigs. The computer-based	
CC	system can identify fragments of the	Enterococcus faecalis genome with	
CC	commercial importance. The products	can be used to detect the presence	
CC	of Enterococcus faecalis in samples.	They can also be used for	
CC	diagnosing Enterococcal infection in	an animal and monitoring	
CC	progression of disease, and for	identifying agents which can be used	
CC	to modulate the growth or pathogenicity	of Enterococcus faecalis, or	
CC	another related organism, in vivo or	in vitro. In particular the	
CC	polypeptides encoded by the	Enterococcus faecalis nucleotide sequences	
CC	can be used in vaccines to prevent or	attenuate an Enterococcal	
CC	infection.		
XX			
SQ	Sequence 6729 BP; 2131 A; 1181 C;	1357 G; 2053 T; 7 other;	
	Query Match	41.3%; Score 565; DB 20; Length 6729;	
	Best Local Similarity	64.5%; Pred. No. 2e-116;	
	Matches 841; Conservative	1; Mismatches 461; Indels	0; Gaps
QY	1	ATGCCATTGTGAAGCTTATCAGACGCTGCAAGCGACGATGCACAAAATGCGTGGTAAG	60
Db	3722	ATGGCTTTTGAAGATTATACAAACCGCCTACGACGCAATGATGTAATAATCCGTGTAAG	3761
QY	61	GGTAAACTTACTGAAGCTGATATAAAGATATATGCTGGAAGTAAAGATTAGCGTTATTT	120
Db	3782	GGAAGAATTCGAGACCGCGACTGAAAAAGAAATGATGCGGAATAATCCGTTTATTA	3841
QY	121	GAGCGTGACGTAANAATCTTTAAAGTTGGTAAAAAGAAATTTATTTAAACACGATGATCAGAA	180
Db	3842	GAGCGCGAGTTAATTTTACAAAGTGTCAAGATTTCACAAAACGTTGTCAGAGACGGCA	3901

Db 521 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATAGCGTTAC 462
QY 119 TTGAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 178
Db 461 TTGAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 402
QY 179 CATTAAGTTCGGATGTAATGCAATCATTAACACCGGGCAACAAAGTTATTAATAATAGTTC 238
Db 401 CATTAAGTTCGGATGTAATGCAATCATTAACACCGGGCAACAAAGTTATTAATAATAGTTC 342
QY 239 AAGATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAATATGTCATAATAAC 298
Db 341 AAGATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAATATGTCATAATAAC 282
QY 299 CACCTACTGTTGTTATGATGGTGGTGTACAGGTGCTGTAAACAAACAACTGCAGGTA 358
Db 281 CACCTACTGTTGTTATGATGGTGGTGTACAGGTGCTGTAAACAAACAACTGCAGGTA 222
QY 359 AATTAGCATTTATGATCGCTAATAAATAACAAACAAACCTATGTTAGTTCGACAGATA 418
Db 221 AATTAGCATTTATGATCGCTAATAAATAACAAACAAACCTATGTTAGTTCGACAGATA 162
QY 419 TTATCGTCCAGCAGCGATATAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 478
Db 161 TTATCGTCCAGCAGCGATATAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 102
QY 479 TATACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGTAACCTAATGCAATTAACCATG 538
Db 101 TATACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGTAACCTAATGCAATTAACCATG 42
QY 539 CTAAAGAGAACATTTAGACTTTGTAATCAATTGATACAGCA 579
Db 41 CTAAAGAGAACATTTAGACTTTGTAATCAATTGATACAGCA 1

RESULT 15
AAS49436/c
ID AAS49436 standard; DNA; 521 BP.
XX
AC AAS49436;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation inhibitory sequence #660.
XX
KW Antisense; ss; prokaryotic cellular proliferation;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 1; Seq ID No 2013; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence is an antisense
CC oligonucleotide of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 521 BP; 139 A; 102 C; 74 G; 206 T; 0 other;

Query Match 37.7%; Score 516.2; DB 23; Length 521;
Best Local Similarity 99.4%; Pred. No. 8.2e-106;
Matches 518; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 59 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATAGCGTTAT 118
Db 521 AGGGTAAACTTACTGAAGCTGATATAAAGATAATGATCGCTGAAGTAAGATAGCGTTAT 462
QY 119 TTGAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 178
Db 461 TTGAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACAGATACAGAACCG 402
QY 179 CATTAAGTTCGGATGTAATGCAATCATTAACACCGGGCAACAAAGTTATTAATAATAGTTC 238
Db 401 CATTAAGTTCGGATGTAATGCAATCATTAACACCGGGCAACAAAGTTATTAATAATAGTTC 342
QY 239 AAGATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAATATGTCATAATAAC 298
Db 341 AAGATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAATATGTCATAATAAC 282
QY 299 CACCTACTGTTGTTATGATGGTGGTGTACAGGTGCTGTAAACAAACAACTGCAGGTA 358
Db 281 CACCTACTGTTGTTATGATGGTGGTGTACAGGTGCTGTAAACAAACAACTGCAGGTA 222
QY 359 AATTAGCATTTATGATCGCTAATAAATAACAAACAAACCTATGTTAGTTCGACAGATA 418
Db 221 AATTAGCATTTATGATCGCTAATAAATAACAAACAAACCTATGTTAGTTCGACAGATA 162
QY 419 TTATCGTCCAGCAGCGATATAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 478
Db 161 TTATCGTCCAGCAGCGATATAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTG 102
QY 479 TATACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGTAACCTAATGCAATTAACCATG 538
Db 101 TATACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGTAACCTAATGCAATTAACCATG 42
QY 539 CTAAAGAGAACATTTAGACTTTGTAATCAATTGATACAGCA 579
Db 41 CTAAAGAGAACATTTAGACTTTGTAATCAATTGATACAGCA 1

Search completed: February 24, 2003, 22:24:30
Job time : 619 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 21:39:11 ; Search time 85 Seconds
(without alignments)
4935.692 Million cell updates/sec

Title: US-09-943-108a-1
Perfect score: 1368
Sequence: 1 atggcatttgaggcttattc.....gtatgaattaccggttttaa 1368

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents.NA.*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1368	100.0	1368	4 US-09-035-382-1	Sequence 1, Appli
2	790.4	57.8	792	4 US-09-035-382-3	Sequence 3, Appli
3	743.2	54.3	1014	4 US-09-134-001C-1958	Sequence 1958, Ap
4	496	36.3	1569	2 US-08-923-772-1	Sequence 1, Appli
5	496	36.3	1569	4 US-09-385-287-1	Sequence 1, Appli
6	496	36.3	7577	4 US-08-961-527-46	Sequence 46, Appli
7	268	19.6	4804	4 US-09-066-047-6	Sequence 6, Appli
8	252.4	18.5	1098	4 US-09-221-017B-928	Sequence 928, App
9	172.6	12.6	836	4 US-08-858-207A-180	Sequence 180, App
10	145.6	10.6	1251	3 US-09-007-476-1	Sequence 1, Appli
11	144.8	10.6	1248	4 US-09-134-001C-1980	Sequence 1980, Ap
12	127.4	9.3	390	4 US-08-651-155B-4	Sequence 4, Appli
13	102.4	7.5	831	3 US-09-007-484-3	Sequence 3, Appli
14	102.4	7.5	831	4 US-09-309-682-3	Sequence 3, Appli
15	102.4	7.5	1278	2 US-08-986-963-1	Sequence 1, Appli
16	102.4	7.5	1278	2 US-08-986-963-3	Sequence 3, Appli
17	102.4	7.5	2764	2 US-08-986-963-4	Sequence 4, Appli
18	100.8	7.4	1293	3 US-09-007-484-1	Sequence 1, Appli
19	100.8	7.4	1293	4 US-09-309-682-1	Sequence 1, Appli
20	100.8	7.4	13121	4 US-08-961-527-126	Sequence 126, App
21	86.8	6.3	1015	4 US-08-981-527A-7	Sequence 7, Appli
22	86.8	6.3	4370	4 US-08-981-527A-20	Sequence 20, Appli
23	61.6	4.5	7218	1 US-08-232-463-14	Sequence 14, Appli
24	61.2	4.5	2166	2 US-08-317-401E-3	Sequence 3, Appli
25	59.6	4.4	2877	2 US-08-317-401E-1	Sequence 1, Appli
26	58	4.2	29793	4 US-09-302-812-38	Sequence 38, Appli
27	58	4.2	29793	4 US-09-511-477-38	Sequence 38, Appli

C	28	58	4.2	29793	4	US-09-511-507-38	Sequence 38, Appl
	29	52.2	3.8	774	4	US-09-461-697-187	Sequence 187, Appl
	30	52.2	3.8	819	4	US-09-461-697-185	Sequence 185, Appl
	31	52.2	3.8	1669	4	US-09-461-697-184	Sequence 184, Appl
	32	49.4	3.6	1022	4	US-08-960-780-33	Sequence 33, Appl
	33	49.4	3.6	1022	4	US-09-073-898-33	Sequence 33, Appl
	34	48.4	3.5	3279	4	US-08-446-137B-1	Sequence 1, Appli
	35	48.2	3.5	3095	6	5231168-1	Patent No. 5231168
	36	47.8	3.5	4074	1	US-08-471-033-19	Sequence 19, Appl
	37	47.8	3.5	4074	2	US-08-471-044-19	Sequence 19, Appl
	38	47.8	3.5	4074	2	US-08-463-483A-19	Sequence 19, Appl
	39	47.8	3.5	4074	2	US-08-471-046A-19	Sequence 19, Appl
	40	47.8	3.5	4074	2	US-08-470-566B-19	Sequence 19, Appl
	41	47.8	3.5	4074	2	US-08-469-334-19	Sequence 19, Appl
	42	47.8	3.5	4074	3	US-09-300-529-19	Sequence 19, Appl
	43	47.4	3.5	717	4	US-09-461-697-189	Sequence 189, Appl
	44	45.4	3.3	1891	4	US-08-973-462-3	Sequence 3, Appli
	45	45.4	3.3	5361	4	US-08-973-462-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-035-382-1
; Sequence 1, Application US/090355382
; Patent No. 6284515
; GENERAL INFORMATION:
; APPLICANT: Black, Michael I.
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES
; TITLE OF INVENTION: AND POLYNUCLEOTIDES
; FILE REFERENCE: GM50035
; CURRENT APPLICATION NUMBER: US/09/035,382
; CURRENT FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: 60/057,890
; EARLIER FILING DATE: 1997-09-03
; NUMBER OF SEQ. ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-035-382-1

Query Match	100.0%	Score 1368;	DB 4;	Length 1368;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1368;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGGCATTGGAAGCTTATCAGAACGCTGTCAGACGCGATGCAAAAATGCGTGGTAAAG	60	
Db	1	ATGGCATTGGAAGCTTATCAGAACGCTGTCAGACGCGATGCAAAAATGCGTGGTAAAG	60	
QY	61	GGTAAACTTACTGAAGCTGATATAAGATAATGCGTGAAGTAAGATTAGCGTTATTT	120	
Db	61	GGTAAACTTACTGAAGCTGATATAAGATAATGCGTGAAGTAAGATTAGCGTTATTT	120	
QY	121	GAGGTGACGTAACTTTAAAGTGGTAAAGATTATTAAACACATGATCAAGACGCCA	180	
Db	121	GAGGTGACGTAACTTTAAAGTGGTAAAGATTATTAAACACATGATCAAGACGCCA	180	
QY	181	TTAGGTTCCGATGTAATGCAATCATTAAACACGCGCAACAAAGTTTAAATAGTTCAA	240	
Db	181	TTAGGTTCCGATGTAATGCAATCATTAAACACGCGCAACAAAGTTTAAATAGTTCAA	240	
QY	241	GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTAAATGTCAAATAACCA	300	
Db	241	GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTAAATGTCAAATAACCA	300	
QY	301	CCTACTGTTTATGATGGTTGTTTACAGGCTGCTGTTAAACAACTCCAGGTAAA	360	
Db	301	CCTACTGTTTATGATGGTTGTTTACAGGCTGCTGTTAAACAACTCCAGGTAAA	360	
QY	361	TTAGCATTATTGCGGTAAATAATACACAAAAAACCCTAATGTTAGTTCACGACGATATT	420	

|||||
Db 361 TTACGATTTATGATCGTAAAAAATACAAAAAACCCTGTTAGTTGCAGCAGATATT 420
QY 421 TATCGTCCAGCAGAGATTAATCAATTACAAACAGTAGGAAACAAATGTGATATTCCTGTA 480
Db 421 TATCGTCCAGCAGCAGATTAATCAATTACAAACAGTAGGAAACAAATGTGATATTCCTGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACAAATGCT 540
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACAAATGCT 540
QY 541 AAAGAAGAACATTTAGACTTTTGTAAATCAATGATTAACAAATGCTTAACAAATGCT 600
Db 541 AAAGAAGAACATTTAGACTTTTGTAAATCAATGATTAACAAATGCTTAACAAATGCT 600
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACAAACGAATTAATGTTA 660
Db 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACAAACGAATTAATGTTA 660
QY 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCATGTTTGCAGATCTTTTGCAGATCAA 720
Db 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCATGTTTGCAGATCTTTTGCAGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
QY 781 TTATCTATTCGTTCCGTCGACACAAACCAATTAATTTTGTGTAAGTGAAGTAA 840
Db 781 TTATCTATTCGTTCCGTCGACACAAACCAATTAATTTTGTGTAAGTGAAGTAA 840
QY 841 GATGGTTTACAGCTATTCATCCCTGAACGCTATGCCATCACGTTATTTAGGTATGGGTGAT 900
Db 841 GATGGTTTACAGCTATTCATCCCTGAACGCTATGCCATCACGTTATTTAGGTATGGGTGAT 900
QY 901 GTGTTAAGTTTAAATGAAAGAGCCCAACAAAGATGTGGATCAAGAAAGAAAGATTTA 960
Db 901 GTGTTAAGTTTAAATGAAAGAGCCCAACAAAGATGTGGATCAAGAAAGAAAGATTTA 960
QY 961 GAGAAAGAGTGGTGGTCACTCGTTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1020
Db 961 GAGAAAGAGTGGTGGTCACTCGTTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 1020
QY 1021 GTGAAATCTAGACCTGATGATGATTAATGAAATGATGAAATGATGAAATGATGAAATG 1080
Db 1021 GTGAAATCTAGACCTGATGATGATTAATGAAATGATGAAATGATGAAATGATGAAATG 1080
QY 1081 AAAGGGCTAGATAAGCTTAATATGAGTGAAGAAAGCAAAATTTGATCAATTAAGCGATATC 1140
Db 1081 AAAGGGCTAGATAAGCTTAATATGAGTGAAGAAAGCAAAATTTGATCAATTAAGCGATATC 1140
QY 1141 CAGTCAATGACGGCTGAAGAAACAAATCCAGACACATTAATGATCACTGATCACTGATCA 1200
Db 1141 CAGTCAATGACGGCTGAAGAAACAAATCCAGACACATTAATGATCACTGATCACTGATCA 1200
QY 1201 CGTATTCCTAAAGGGTCGTCGTTTCATTAAGAAAGTCAATCGTTTGCATGAAACAAATTT 1260
Db 1201 CGTATTCCTAAAGGGTCGTCGTTTCATTAAGAAAGTCAATCGTTTGCATGAAACAAATTT 1260
QY 1261 AACGATATGAAGAAATGATGAACAAATTCACCTGTTGGGGGTAAGGTAAGAAAGGTAAG 1320
Db 1261 AACGATATGAAGAAATGATGAACAAATTCACCTGTTGGGGGTAAGGTAAGAAAGGTAAG 1320
QY 1321 GCATCAATCAATGAAATATGTTAAAGGTATGAATTTACCGTTTAA 1368
Db 1321 GCATCAATGAAATATGTTAAAGGTATGAATTTACCGTTTAA 1368

RESULT 2
US-09-035-382-3
; Sequence 3, Application US/09035382
; Patent No. 6284515
; GENERAL INFORMATION:

; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES
; FILE REFERENCE: GM50035
; CURRENT APPLICATION NUMBER: US/09/035,382
; CURRENT FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: 60/057,890
; EARLIER FILING DATE: 1997-09-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-035-382-3

Query Match 57.88; Score 790.4; DB 4; Length 792;
Best Local Similarity 99.94; Pred. No. 6.2e-191;
Matches 791; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGGCTTTATCAGAACGCTTCAAGCGACGATGCAAAAAATGCGTGTAAAG 60
Db 1 ATGGCATTTGAAGGCTTTATCAGAACGCTTCAAGCGACGATGCAAAAAATGCGTGTAAAG 60
QY 61 GGTAAACTTACTGAAGCTGATTAAGATTAATGATCGTGAAGTAAGATAGCGTTATTT 120
Db 61 GGTAAACTTACTGAAGCTGATTAAGATTAATGATCGTGAAGTAAGATAGCGTTATTT 120
QY 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
Db 121 GAGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
QY 181 TTAGTTCGCGATGTAATCAATTAACACAGCGGCAACAAAGTTATTAATAATAGTTCAA 240
Db 181 TTAGTTCGCGATGTAATCAATTAACACAGCGGCAACAAAGTTATTAATAATAGTTCAA 240
QY 241 GATGAATTAACAGAGTTGATGGTGGAGAAATFACATCGATTAATATGTCATTAACCA 300
Db 241 GATGAATTAACAGAGTTGATGGTGGAGAAATFACATCGATTAATATGTCATTAACCA 300
QY 301 CCTACTGTTTATGATGTTGGTTTACAGGTGCTGTTAAACCAACACGTCAGGTAAA 360
Db 301 CCTACTGTTTATGATGTTGGTTTACAGGTGCTGTTAAACCAACACGTCAGGTAAA 360
QY 361 TTAGCATTTATGATGCTGTTAAATATCAACAAAAACCTATGTTAGTTGCAGCAGATTT 420
Db 361 TTAGCATTTATGATGCTGTTAAATATCAACAAAAACCTATGTTAGTTGCAGCAGATTT 420
QY 421 TATCGTCCAGCAGCGATTAATCAATTAACACAGTAGGAAACAAATTTGATATTCCTGTA 480
Db 421 TATCGTCCAGCAGCGATTAATCAATTAACACAGTAGGAAACAAATTTGATATTCCTGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACAAATGCT 540
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATCAATGATTAACAAATGCT 540
QY 541 AAAGAAGAACATTTAGACTTTTGTAAATCAATGATTAACAAATGCTTAACAAATGCT 600
Db 541 AAAGAAGAACATTTAGACTTTTGTAAATCAATGATTAACAAATGCTTAACAAATGCT 600
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACAAACGAATTAATGTTA 660
Db 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAACAAACGAATTAATGTTA 660
QY 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCATGTTTGCAGATCTTTTGCAGATCAA 720
Db 661 GTTGTGATTAATCAATGACGGGTCAAGATGCTGTCATGTTTGCAGATCTTTTGCAGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA 780
QY 781 TTATCTATTCGTTCCGTCGACACAAACCAATTAATTTTGTGTAAGTGAAGTAA 792


```
QY 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAG 60
Db 1 ATGGCAATTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAG 60
QY 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTATTT 120
Db 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTATTT 120
QY 121 GAGGCTGACGTAACCTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGA 180
Db 121 GAGGCTGACGTAACCTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACGCGA 180
QY 181 TTAGGTTCCGAGTAAGTCAATCAATTAACACAGGCGCAACAGTATTAATAATAGTCAA 240
Db 181 TTAGGTTCCGAGTAAGTCAATCAATTAACACAGGCGCAACAGTATTAATAATAGTCAA 240
QY 241 GATGAATTAACGAGTGTAGGCGTGAGAGAAATATCATGATTAATATGTCATAAACA 300
Db 241 GATGAATTAACGAGTGTAGGCGTGAGAGAAATATCATGATTAATATGTCATAAACA 300
QY 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGTTAAACCAACAACTGCAGGTAA 360
Db 301 CCTACTGTTGTTATGATGTTGTTTACAGGTGCTGTTAAACCAACAACTGCAGGTAA 360
QY 361 TTAGCATTATGATGCGTAAATAACAAACAAACAAACCTGTTAGTTCAGACAGATTT 420
Db 361 TTAGCATTATGATGCGTAAATAACAAACAAACAAACCTGTTAGTTCAGACAGATTT 420
QY 421 TATCGTCCAGAGCGATTAATCAATTAACAAACAGTAGGAAACAAATGATATTCCTGTA 480
Db 421 TATCGTCCAGAGCGATTAATCAATTAACAAACAGTAGGAAACAAATGATATTCCTGTA 480
QY 481 TACAGTGAAGAGATCAAGTAAGACCAACAAATTTAACTAATGATCAATTAACAAATGCT 540
Db 481 TACAGTGAAGAGATCAAGTAAGACCAACAAATTTAACTAATGATCAATTAACAAATGCT 540
QY 541 AAAGAAACAACTTTAGATTTGTAATCAATGATACAGAGGTGCGATTACACATCGATGAA 600
Db 541 AAAGAAACAACTTTAGATTTGTAATCAATGATACAGAGGTGCGATTACACATCGATGAA 600
QY 601 GCATTGATGAAGCAATTAAGAGAGTAAAGAAATTTCTAAACCAACAGAAATATGTTA 660
Db 601 GCATTGATGAAGCAATTAAGAGAGTAAAGAAATTTCTAAACCAACAGAAATATGTTA 660
QY 661 GTTGTGATTAATCAAGCGGTCAAGATGCTCAATGTTGAGAACTCTTTGACGATCAA 720
Db 661 GTTGTGATTAATCAAGCGGTCAAGATGCTCAATGTTGAGAACTCTTTGACGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGTTGATACACGTTGTTGTCAGCT 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACATAATAGATGTTGATACACGTTGTTGTCAGCT 780
QY 781 TTATCTATTCGTTGCGTGACACAAACCAATTAATTTGTTGTTGATGAGTGAAGATTA 840
Db 781 TTATCTATTCGTTGCGTGACACAAACCAATTAATTTGTTGTTGATGAGTGAAGATTA 840
QY 841 GATGTTTATGAGCTATCCATCCGTAACGATGCTATCACTATTTAGTATGTTGGTGTAT 900
Db 841 GATGTTTATGAGCTATCCATCCGTAACGATGCTATCACTATTTAGTATGTTGGTGTAT 900
QY 901 GTGTTAAGTTAATTGAAGAAACGCAACAGATGTTGATCAAGAAAGAAAGATTTA 960
Db 901 GTGTTAAGTTAATTGAAGAAACGCAACAGATGTTGATCAAGAAAGAAAGATTTA 960
QY 961 GAGAAAGATGCGGTGAGTCACTGTTTACCTTTAGATGATTTTTTGAACAACTGATCAG 1020
Db 961 GAGAAAGATGCGGTGAGTCACTGTTTACCTTTAGATGATTTTTTGAACAACTGATCAG 1020
QY 1021 GTGAAAGATGAGGCGGATGGAAGCTTGTCTCAAGATGATTTCCAGGATGAGGCAACAT 1080
Db 1021 GTGAAAGATGAGGCGGATGGAAGCTTGTCTCAAGATGATTTCCAGGATGAGGCAACAT 1080
QY 1081 AAAGGGCTAGATAAGCTTAATATGATGAGTGAAGAAAGCAAAATTTGATCATATTAAAGCGATTATC 1140
```

```
Db 1081 CCAGCACTTAAACATGAAGTGGATGAACCCAGATTTGCTGTAAACGTGCCATTGTG 1140
QY 1141 CAGTCAATGAGCGCGCTGGAAGAAACAATCCAGACACATGGAATGATCACTCACTAAAAG 1200
Db 1141 TCTTCATGACATCUGAAGAACGTAAGAACCCAGATTTGTTAAATCCAGCCGCTGCCGT 1200
QY 1201 CGTATTGCTAAAGGCTGCTGCTTCACTTACAGAAAGTCAATCTTTGATGAACAATTT 1260
Db 1201 CGTATTGCTGCTGCTGCTGGAATACATCTGCGAAGTCAATAAATTCATCAAGGACTTT 1260
QY 1261 AAGATATGAGAAATGATGAGAAACAATTCATCTGCTGCGGCTAAAGTAAAA 1312
Db 1261 AACCAGCTTAAACAGCTCATCGAGGCTGTATGCTCTGGGATATGAATAAAA 1312
```

RESULT 5

```
US-09-385-287-1
; Sequence 1, Application US/09385287
; Patent No. 6350857
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL fih
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GN10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-385-287-1
```

Query Match 36.3%; Score 496; DB 4; Length 1569;
Best Local Similarity 61.1%; Pred. No. 2.3e-116;
Matches 802; Conservative 0; Mismatches 510; Indels 0; Gaps 0;

```
QY 1 ATGCCATTTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAG 60
Db 1 ATGCCATTTGAGGCTTATCAGAACCGCTGCAAGCGAGATGCAAAAATGCGTGTAG 60
QY 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTATTT 120
Db 61 GGTAACTTACTGAGCTGATATAAGATAATGATGCGTGAAGTAAGATAGCGTATTT 120
QY 121 GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTTATTAACAACTATCAGAACGCGCA 180
```

[illegible]

Db 1201 CQTATGCTGCTGGTTCGTGGAATACATCTTCGGAAGTCATATAATATCATCAAGGACTTT 1210
QY 1261 AACGATATGAAGAAAATGATGAACAAATCACTGGTGGCGGTAAAGGTAAAA 1312
Db 1261 ACCAGGCTAAACAGCTCATCCAGGCTTATGCTGGGATATGAATAAAA 1312
RESULT 6
US-08-961-527-46/c
; Sequence 46, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-46

Query Match	36.3%;	Score 496;	DB 4;	Length 7577;
Best Local Similarity	61.1%;	0;	Mism. 4.2e-116;	
Matches 802;	Conservative	0;	Indels	Gaps 0;
QY 1	ATGCGATTGTGAAGCTTATCAGAACGCCCTGCAGGACGATGC	AAAAATCGCTGGTGAAG	60	
Db 3033	ATGCGATTGTGAAGCTTTAAACAGAACCTTTGCAGAACCTCTTTAA	AAATCTACGCTAAAAAA	2974	
QY 61	GGTAAACTTACTGACGTGATATAAAGATATATGCTGAC	TGACATACGCTTATTT	120	
Db 2973	GGAAATATCTCTGAATCTGATGTCCAGAGCGACCAAGAAATTCGCTT	GGCCCTTGCTC	2914	
QY 121	GAGGTGCAGTAAACTWTTAAAGTGGTAAAAAGAAATTTATTA	AAACAGATCATGAACGGCGCA	180	
Db 2913	GAGGCCGAGCTTGCCTTGTGTAAAGACCTTTATCAGA	AAATTCGCTGACCGTGCA	2854	
QY 181	TTAGGTTCOGATGTATPGCAATCATTAACACCGGCAACAGT	TATTAATAATATTCAA	240	
Db 2853	GTGCGGCATGAGGTCATTGTATACACTTAATCCTCGCAACAGAT	TATTAATAATCGTTGAT	2794	
QY 241	GATGAATTAACGAAGTTGATGGTGGAGAAATACATCG	TATTAATGTCTAAATAACCA	300	
Db 2793	GAGGAATTCGACGCGGTTTTAGTGTCTGATACGGCAAGAAATAT	CAGCTACCTTAAGATT	2734	

QY	301	CTTACTGTTGTTATGATGGTGTGGTTTACAAGSGTCTGCTGTAAACAACAACACTGCAGGTAAA	360
Db	2733	CCAAACATCATCATGATGGTGTGGTTTACAAGGGGCTGGTAAACAACACTTTCGTCTGTAAA	2874
QY	361	TTAGCATTTATTGATGCGGTAAAAAATACAACAACAAAACCCATGCTTAGTTTGCAGCAGATATT	420
Db	2673	TTGGCCACAACAACACTCAAGAAAGAGAAAATGCTGTCTCTTTGATGATTCGGCGGGATATT	2614
QY	421	TATCGTCCAGCAGGATAAATCAATTTACAACAACAGTAGAGGAAACAACAAATGATATTCCTGTAA	480
Db	2613	TATCGTCCAGCTGCCATTGACCAGCTTAAGACCTTTGGGACACACAGATTGATGTGCTCTGC	2554
QY	481	TACAGTGAAGAGAGATCAAGTAAAGCCACAACAATAATGTGTACTATGCAATTAACAACATGCT	540
Db	2553	TTTGCACTTTGGAAACAGAGTACAGACTGTTGAGATTGTAGTGTGAGTCAAGGTTTTGGAGCAAGCC	2494
QY	541	AAAGAAGAACATTTAGACTTTGTATCAATTCATACAGCAGGTCGATTACACATPCATGACAA	600
Db	2493	CAAACTAACTAACAGACTATGCTTTGATTGATACCTTGGGGTGTGTTGCGAGATTGATGAG	2434
QY	601	GCATGATGAAGCAATTAAGAAGTAAAGAATAATGCTTAAACAAACAGAAATATGTTTA	660
Db	2433	CTCCTCATGAATGAGCTTCTGTAITGTAAGAACATTTGGCTCAACCAAAATGAAATCTTGCTT	2374
QY	661	GTTGTGATTAAGTACGGGTCAGAGTCTGTCAATGTTGCAGAACTCTTTTGCAGNCAAA	720
Db	2373	GTCGTTGATGCTATGATTGTGTGAGAGCAGCCAAATGTTGGCGTGAATTTAATGCTCAG	2314
QY	721	CTTGATGTCACAGSGTGTACCTTTAATATAATAGATGGTGATACACGTGGTGGTGCAGCT	780
Db	2313	TTGGAAGTGACTGGGGTCATCCTTACCAAGATTGATGGGGAATCTCGTGGTGGTCTGCT	2254
QY	781	TTATCTATTTCGTTGCGGTGACACAAAACCAATTAATTTGTTGGTATGAGTGAAGAATTA	840
Db	2253	CTGTCTGTTGTCACATTACTGTGAACAAACCAATCAAGTTCACCTGCTACAGGTGAAGAAGAT	2194
QY	841	GATGGTTAGACCTATTCCTCATCTGCAAGSTATGTCATCAGTATTTTATGCTATGSGTGAT	900
Db	2193	ACGGACATTTGAACCTTCCACCCAGCAGCGATGTCTAGCCGTATCCCTTGTATGSGGGGAT	2134
QY	901	GTGTTAAGTTAATTTAAAAGCGCAACAAGATGTGGATCAAGAAAAGCAAAAAGATTTA	960
Db	2133	ATGCTCACCTTGATTGAGAAGCTTCTCAGGAATACGATGAACAAAAGCCCTTGAATG	2074
QY	961	GAGAAAAAGATCGCTGAGTCATCGTTTACTTTAGATGATTTTTTGAACAACACTTGATCAG	1020
Db	2073	CTCTGAGAAGATCGCGAAAACACCTTTGATTTTAATGATTTCAWCGATCAATTTAGATCAG	2014
QY	1021	GTGAAAAATCTAGGACCACATGGATGATATATGAAAATGATTCACAGGTATGTAATAAATG	1080
Db	2013	GTGCAAAATATGGCGCGTGGAGAGCTTGCTCAGATGATTTCCAGGTATGSGCCACAAAT	1954
QY	1081	AAAGGGCTAGATAGCTTTAATATGATGTAAGAACAATATGATCATPATAAAGCGATTATC	1140
Db	1953	CCAGCCCTTCAAAAACATGAAGSGTGAAGACCCAGATTGCTGCTGTAAACGCTGCCATTGTG	1894
QY	1141	CAGTCAATGACCGCGCTGAAGAAACAAATCCAGACACTTGAATGTATCACGTAAAAAG	1200
Db	1893	TCTTCGATGACACCTTGAGAGCGTGAACACCCAGATTGTTTAATCCARCGCCTGCGCCT	1834
QY	1201	GCTATTTGCTAAAGGGCTTGGTCTGCTTCAATTAACAAGAAGTCAATCGTTTGATGAACAATTT	1260
Db	1833	CGTATTCGCTGGTCTGGAATACATTCGTCGAAGTCAATTAATTCATCAGGACTTT	1774
QY	1261	AACGATNGAAGAAATGANGAACAAATTCATCTGCTGGCGGTAAAGSTAAA	1312
Db	1773	AAACAGGCTAAACAGCTCATCGAGGCTGTTATGCTCGGGGATATCAATPAAA	1722

RESULT 7

RESOL /
ITS-09-066-047-6

US-09-060-047-0
; Sequence 6, Application US/09066047A; sequence 0, APPEND
; Patent No. 6306394

```
QY 367 TTATTGATCGGTAAATAACACAAAAAACCCTATGTGTTGTCGACGACAGATATTATCGT 426
Db 3855 --ACTGAGTTAAGAAGAGATCTTAAACCCGTTGGTAGCGTCTTTAGACGATATATCGT 3912
QY 427 CCACGACGATAAATCAATTAACAACACAGTAGGGAACAAATTCATATTCTCTGTATACAGT 486
Db 3913 COTGACGCTCGAGACAGCTGAAGCTTTTGCTGATGGGTTGGTATAGACAGCTTCCC 3972
QY 487 GAAGAGATCAAGTAAGACCAACAATAATTTCTAATCAATCAATTAACACATGCTTAAGAA 546
Db 3973 APTCTTGAGGACAAACCACTTGATATTGCGAAGCGTCTATGAGGGAAGCGAGGCTC 4032
QY 547 GAACATTTAGACTTTGTAATCATATGATACAGCAGCTGATACACATGATGACATGTC 606
Db 4033 AAAGGACACAGTGTGTGTTTGGATACAGCGGGCGCTTGATATCAATCAGACATG 4092
QY 607 ATGACGAATTAAGAAGTAAGAAATTCCTAAACCAACGAATTTATGTTATGTC 666
Db 4093 ATAGATGAGCTGAAGTGTGAAGAAGGAGGTATCACACGCTGAAATTTGTTGTTGA 4152
QY 667 GATTCATGACGGGTCAAGTCTGTCAATGTTCCAGAAATCTTTGACATCAACTTGAT 726
Db 4153 GACTCCTTAATGGGCAAGATGCCCTCACTATGTCGCAAGTCAATGAGGAGTTAGGC 4212
QY 727 GTCACAGGTGTACCTTAACATAATAGATGCTGATACAGCTGGTGGTGCACGCTTATCT 786
Db 4213 ATTACTGGGACGATCTTTACAGGGCGATGCTGATCCTAGGGGTGGTCTATCTTGCT 4272
QY 787 APTCGTTCGGTACACAAAACCAATTAATTTGTTGATGAGTGAAAGTTAGATGGT 846
Db 4273 ATGAAGTTGGTGTCTGGATGCTCTATAAAGTTCATGCTACGGGAGAGAGCTGAGAT 4332
QY 847 TTAGACTATTCATCTCTGAACTGATGCGATCACTATTTAGTAGTGGTGTATGTTGA 906
Db 4333 TTGACGATTTCTATCTCTGATAGATAGCTCTGTAATGTTAAATATGGAGATCTGCA 4392
QY 907 AGTTTAATGAAGAAGCGCAACAGATCTGGATCAAGAAACAAAGAAATTTAGAGAA 966
Db 4393 TCTCTTTGAGAAAGCGGTAGACGGGTGGCAAGGATACAAATTAATGAGCTACAGCG 4452
QY 967 AAGATGCTGATGATCTGTTTACTTTAGATGATTTTTAGAACAACTTGATCAGTGAAA 1026
Db 4453 AAGCCAGAGAGGTAAATTCGATTTGGATGATCTTTTATTCAGCTGAACCTTTGAT 4512
QY 1027 AATCTAGGACCACTGGATGATATTATGAATGATCCAGGTAT---GATATAAATGAAA 1083
Db 4513 AAAATGGGTGTATGCTAATATAATGAAGTTTATACCGCTTTGGTAAAGATATAAA 4572
QY 1084 GGGCTAGATAGCTTAATATGATGTAAGCAAAATTCATATTAACCGATATCCAG 1143
Db 4573 CCAAGTTGGGGGATAGCTGATGACAGCAAAAGTCAGATGATACATTCGATTTAAAC 4632
QY 1144 TCAATGACCGCGGTGAAAAGAACATCCAGACACATTTGATATCACTGATTAAGAGCGT 1203
Db 4633 TCAATGACGAGGAGGAGGCGGATCTCTGATGATCTGATGATGCTGCGGAGGAGCAAG 4692
QY 1204 APTGCTAAAGGCTCTGCTGTTCAATACAGAGATCAATCGTTTGATGAAACAAATTAAC 1263
Db 4693 APAGCAAGGTCGGGAGCTAAGGTTGATGCTGTAATCGCTTGTGTAAGAGCATAT 4752
QY 1264 GATATGAA 1271
Db 4753 CAGATGAA 4760
```

RESULT 8

```
US-09-221-017B-928/G
; Sequence 928, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
```

```
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 06/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 928:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: 1...1098
; US-09-221-017B-928
```

```
Query Match 18.5%; Score 252.4; DB 4; Length 1098;
Best Local Similarity 53.6%; Pred. No. 8.2e-55;
Matches 548; Conservative 0; Mismatches 471; Indels 3; Gaps 1;

QY 55 GGTAAAGGTAACTTGTAGTGAAGCTGATATAAAGATAATGATGCTGGAAGTAAGATTACG 114
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1092 GGTGAGGGGAGAAATTACCGAGATCAATGTAGCCGAAACTCTCAAGATGTACGTCGTGCC 1033

QY 115 TTATTGAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACATATCAGAA 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1032 CTTTGGATGCCGACGTAATCTGCTGTGCCAAGCAGTTTACGACCTTGGTGAAGGAA 973

QY 175 CGCGATTAGTTCGGATGTAATCAATCAATACACAGGCGCAACAACTTATTATAATA 234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 AAAGCCATCGGCCAAACACGTGCTGACCTCCGTCGCGGCAACTGATGCTGAAGATC 913

QY 235 GTTCAAGATGAATTAACAGATTGATGGTGGAGAAATACATCGATTATATGTCAAAT 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 GTTCATGACGAATTTGGTCTGCTGTGATGGCGACATCTCTGTGGATATTAACTCT---CAA 856
```

```
QY 295 AAACACCTACTGTGTGTTATGATGGTTGTTTACAGGTGCTGTAACAAACAACTGCA 354
    || || || || || || || || || || || || || || || || || || || || ||
Db 855 GGCAGTCCGGCAGTAATCTCGATCCGATTCGAGTTCGGTGAAGACCACTCTCTCC 796
    || || || || || || || || || || || || || || || || || || || || ||
QY 355 GGTAAATAGCATATATGATGCGGTAAATAATACAAACAAACCTAATGTTAGTTCAGCA 414
    || || || || || || || || || || || || || || || || || || || || ||
Db 795 GGCAAACTGGCCAAATATGCTAGTCCAAAGAGGCAACACCTCTGCTGCTGCTGCTGC 736
    || || || || || || || || || || || || || || || || || || || || ||
QY 415 GATATTTATCTCCACAGCAGTAAATCAATTACAAACAGTAGGAAACAAATGATAT 474
    || || || || || || || || || || || || || || || || || || || || ||
Db 735 GAGGTGTATCGCGCTCGCGTATCCAGCAGCTGATGATCCCGCGAGCAGCTCGGGTA 676
    || || || || || || || || || || || || || || || || || || || || ||
QY 475 CCGTATACAGTGAAGAGATCAAGTAAGAACCCACAAACAAATGTAATGATCAATTAATA 534
    || || || || || || || || || || || || || || || || || || || || ||
Db 675 CCGGTCTACAGCAGCAGCAGCAAAAGCGGTGAGATAGCCGACACGCAATAGCC 616
    || || || || || || || || || || || || || || || || || || || || ||
QY 535 CATGCTAAAGAGAAACATTTAGACTTTGTAATCATGATACAGCAGTGCATTAACATC 594
    || || || || || || || || || || || || || || || || || || || || ||
Db 615 GAAGCCAAAGACCAAGGGAACGAGTTGTATCATGACACCGCGGTCTGTTGGCCATC 556
    || || || || || || || || || || || || || || || || || || || || ||
QY 595 GATGAAGCATTTGATGAACGAAATTAAGAAAGTAAAGAAATTTGCTAAACCAACAGAAAT 654
    || || || || || || || || || || || || || || || || || || || || ||
Db 555 GACGAAGAAATGATCGCGCAGATAGAGGCCATCAGGCTGCCATCCAGCCGACGAGATC 496
    || || || || || || || || || || || || || || || || || || || || ||
QY 655 ATGTTAGTTGTCGATTCATGAAGGTCAGATGCTGTCAATGTTCAGAAATCTTTTGAC 714
    || || || || || || || || || || || || || || || || || || || || ||
Db 495 CTTTTCGTGTTGACTCGATGACCGGTGACGAGCGCGTGAATACAGCAAGGAGTTCAAT 436
    || || || || || || || || || || || || || || || || || || || || ||
QY 715 GATCAACTGTGATCAGAGTGTACCTTAACATAATAGATGGTGATACAGTGTGTGT 774
    || || || || || || || || || || || || || || || || || || || || ||
Db 435 ACACGCTTGGATTTGACGCGTGTGTCTTTACCAAGCTCGATGAGATACCGGGCGGT 376
    || || || || || || || || || || || || || || || || || || || || ||
QY 775 GCAGCTTTATCATTTCTGTTGTCGACACAAAAACCAATTAATTTGTTGATCACTGAA 834
    || || || || || || || || || || || || || || || || || || || || ||
Db 375 CCGCCCTCTCGATCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 316
    || || || || || || || || || || || || || || || || || || || || ||
QY 835 AAGTTAGATGGTTTATGAGCTATTCATCCATGACGATGATGATGATGATGATGATGAT 894
    || || || || || || || || || || || || || || || || || || || || ||
Db 315 AAGATGAAGCAATCGATGCTTTTCAACCGGAGCGTATGGCGACCGCTTCCTTGGCATG 256
    || || || || || || || || || || || || || || || || || || || || ||
QY 895 GGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 954
    || || || || || || || || || || || || || || || || || || || || ||
Db 255 GGGGACATCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 196
    || || || || || || || || || || || || || || || || || || || || ||
QY 955 GATTTAGAAAAAGATCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1014
    || || || || || || || || || || || || || || || || || || || || ||
Db 195 AAAGTGGAGAGAAATAGCCAGAAATCAGTTCGACTTCAATGACTTCCTTGGCCAGATA 136
    || || || || || || || || || || || || || || || || || || || || ||
QY 1015 GATCAGTGAATAACTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGAT 1074
    || || || || || || || || || || || || || || || || || || || || ||
Db 135 CATCAGATCAAAAAGATGGTAACTCAAAAGCTTGATCATCATGATCCCGGAGTGGGA 76
    || || || || || || || || || || || || || || || || || || || || ||
QY 1075 AA 1076
    ||
Db 75 AA 74
```

RESULT 9

US-08-858-207A-180

; Sequence 180, Application US/0858207A

; Patent No. 6348328

; GENERAL INFORMATION:

; APPLICANT: Black, Michael

; APPLICANT: Hodgson, John

; APPLICANT: Knowles, David

; APPLICANT: Nicholas, Richard

; APPLICANT: Stodola, Robert

; TITLE OF INVENTION: No. 6348328el Compounds

; NUMBER OF SEQUENCES: 552

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

```
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmli, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 180:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-858-207A-180
```

Query Match 12.6%; Score 172.6; DB 4; Length 836;

Best Local Similarity 61.9%; Pred. No. 1.1e-34;

Matches 274; Conservativity 0; Mismatches 169; Indels 0; Gaps 0;

```
QY 1 ATGCGATTTGAAGGCTTTATCAGACGCTCCACAGCGCATGCAACCAAAATCCGTGGTAA 60
    || || || || || || || || || || || || || || || || || || || || ||
Db 393 ATGCGATTTGAAAGTTTAAACAGACGTTTCAGACGCTCTTTAAAAATCTACGTA AAAA 452
    || || || || || || || || || || || || || || || || || || || || ||
QY 61 GGTAAACTTACTGAAGCTGATATAAAGATAATGATCGTGAAGTAAGATTAGCTTATTT 120
    || || || || || || || || || || || || || || || || || || || || ||
Db 453 GGAATAATCTCTGATCTGATGTCCAGAGGCAACCAAGAAATTCGCTGGCGCTGCTC 512
    || || || || || || || || || || || || || || || || || || || || ||
QY 121 GAGCTGAGCTAAACTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAAGACGCGCA 180
    || || || || || || || || || || || || || || || || || || || || ||
Db 513 GAGCCGACGCTTGCCTTGTAAAGGACTTTATCAGAAAAGTTCGTGAGCGCGCA 572
    || || || || || || || || || || || || || || || || || || || || ||
QY 181 TTAGTTCGCGATGTAATGCAATCATTAACACGAGGCGCAACAGTTATTAAATAAGTCAA 240
    || || || || || || || || || || || || || || || || || || || || ||
Db 573 GTGCGGATGAGTCAATGATACACTTAATCCCTGCGCAACAGATTATTAAATCGTTGAT 632
    || || || || || || || || || || || || || || || || || || || || ||
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTAAATATGTCAAATAAACCA 300
    || || || || || || || || || || || || || || || || || || || || ||
Db 633 GAGAACTGACACCGCTTTTAGTTCGTGATACGCGAGAATTTATCAAGTCACCTAAGATT 692
    || || || || || || || || || || || || || || || || || || || || ||
QY 301 CCTACTGTTGATGATGGTTTACAGGTTGCTGGTAAACCAACAACTGCGAGGTAAA 360
    || || || || || || || || || || || || || || || || || || || || ||
Db 693 CCAACCATCATGATGATGGTTGTTTACAGGGGCTGTTAAACCAACCTTTGCTGTGAAA 752
    || || || || || || || || || || || || || || || || || || || || ||
QY 361 TTACATATTGATCGTAAATAATACAAACAAAACCTATCTTAGTTGCGAGCATATT 420
    || || || || || || || || || || || || || || || || || || || || ||
Db 753 TTGCCAACAACTCAAGAAAGAAAATGCTCGTCTTTGATGATTGCGGCGGATATT 812
    || || || || || || || || || || || || || || || || || || || || ||
QY 421 TATGTCACGACGCGATAATCA 443
    || || || || || || || || || || || || || || || || || || || || ||
Db 813 TATGTCACGCTGCCATTGACCA 835
    || || || || || || || || || || || || || || || || || || || || ||
```

RESULT 10

US-09-007-476-1

; Sequence 1, Application US/09007476

; Patent No. 6159949

; GENERAL INFORMATION:

APPLICANT: Black, Michael T.
TITLE OF INVENTION: No. 6159949el Ptsy
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/007,476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM10079
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1251 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-007-476-1

Query Match 10.6%; Score 145.6; DB 3; Length 1251;
Best Local Similarity 53.6%; Pred. No. 8.9e-28;
Matches 388; Conservative 0; Mismatches 309; Indels 27; Gaps 3;

QY 276 ATCGATTAAATGTCAAAATAAACCACCTACTGTGTTATGATGGTTGGTTTACAGAGTGC 335
Db 531 ACCTATGACATAGACATGGTCGTTTAAATGCTCATTTTAAGTCGGTGTGAATGGTGT 590
QY 336 TGGTAAACAACTGCGAGGTAATATAGCATTTATGATCGGTAAATAATACAAACAAA 395
Db 591 TGGTAAACAACTGCGAGGTAATATAGCATTTATGATCGGTAAATAATACAAACAAA 647
QY 396 ACCTATGTTAGTGCAGCAGATTTATCGTCCAGCAGCATTAATCAATTAACAACAGT 455
Db 648 AGTAATGTTAGTGCAGCAGATTTATCGTCCAGCAGCATTAATCAATTAACAACAGT 707
QY 456 AGGAAACAAATGATATTCCTGTATACAGTGAAGGAGATCAAGTAAAGCCACACAAAT 515
Db 708 GSGCGAAGCTGTTGGTGTAGATGTAATAGCCAAAGTGAAGTCTGTGATCCAGTGTGT 767
QY 516 TGTAACTAATGCAATTAACATGCTTAAAGAGACATTTAGATCTTTTAATCATTTGATAC 575
Db 768 TATGATGATCGGATTAATGCGCTTAAACAAAGGTTGATATTTTAAATCTGTGATAC 827
QY 576 ACCAGTGCATACATCGATCAAGCATTTGATGAACGATTAAGAGAGTAAATGATGATG 629
Db 828 CGCTGGAGGTTTAAACAAATTAACAAATTAATGCAAGATTAAGAAAGTGAATGCGGTGT 887
QY 630 -----AGAAATGCTAAACAAACGAAATATTTAGTTGTGCGATCAATGAC 677
Db 888 AATTATCGACATGCGCAGTGCCTCATGGAAGCATTTACTATGTTTATGATGCTACAC 947
QY 678 GGGTCAAGATGCTGTAATGTTGCAAGATCTTTTGAAGATCAACTTTGATGTCACAGTGT 737
Db 948 TGGTCAAGATGCTGTAATGTTGCAAGATCTTTTGAAGATCAACTTTTGAAGATGTTTACAGTAT 1007

QY 738 TACCTTAACATAATAGATGGTGATACACGTTGGTGGTGAGCTTTATCTATTCGTTGGT 797
Db 1008 TGTATTAAACGAAATAGATGGTGATACACGCAAGGTGGTGTATTAGCCATTCGTAATGA 1067
QY 798 GACACAAAACCAATTAATTTGTTGGTATGAGTGAAGAAAGTTAGATGGTTAGAGCTATT 857
Db 1068 AWTGCACATTCAGATTAAATATGTTAGGTTTAGTGAGCAATTAGATGACTTACAAACATT 1127
QY 858 CCATCCTGAACGATGGCATCACGCTATTTTAGGTATGGTGATGGTGTATTAAGTTTAATGA 917
Db 1128 TAACTCTGAAGGTTATGCTACGGCTTAT-----TCGCTGATATGATTGAACAATGA 1181
QY 918 AAAAGCGCAACAAGATGTTGATCAGAGAAAAGCAAAAGATTTAGAGAAAAGATGCGTGA 977
Db 1182 CGAAATAACAACAGATTGAAATGATCAAAATGTAACAGAGAAAAGGACGATATCATGG 1241
QY 978 GTCA 981
Db 1242 GGCA 1245

RESULT 11

US-09-134-001C-1980
; Sequence 1980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1980
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1980

Query Match 10.6%; Score 144.8; DB 4; Length 1248;
Best Local Similarity 54.6%; Pred. No. 1.4e-27;
Matches 346; Conservative 0; Mismatches 267; Indels 21; Gaps 2;

QY 276 ATCGATTAAATGTCAAAATAAACCACCTACTGTGTTATGATGGTTGGTTTACAGAGTGC 335
Db 540 AGCAATGATATGAGATGGAGCTTTTAAATGTCATAGTGTGTTGGTGTGATGGTGT 599
QY 336 TGGTAAACAACTGCGAGGTAATATAGCATTTATGATCGGTAAATAATACAAACAAA 395
Db 600 CGGCAAAACAACTGCGAGGTAATATAGCATTTATGATCGGTAAATAATACAAACAAA 656
QY 396 ACCTATGTTAGTGCAGCAGATTTATCGTCCAGCAGCATTAATCAATTAACAACAGT 455
Db 657 AGTAATGTTAGTGCAGCAGATTTATCGTCCAGCAGCATTAATCAATTAACAACAGT 716
QY 456 AGGAAACAAATGATATTCCTGTATACAGTGAAGGAGATCAAGTAAAGCCACACAAAT 515
Db 717 GSGCAGAGCTGTTGGTGTGAGTGTGAGTCAAAAGAGGTTCTGACCCCTGCAGCAGT 776
QY 516 TGTAACTAATGCAATTAACATGCTTAAAGAGACATTTAGATCTTTTAATCATTTGATAC 575
Db 777 AGTATGATGCGATTAATGCGCAAAATTAAGGCGTATGATTTTAAATTTGATGATAC 836
QY 576 ACCAGTGCATTAACATCGATGAGACATTTGATGAACGATTAAGAGAGTAAAGAGAAAT 635
Db 837 TCGAGGACGCTTGAATTAATCTAATTAATGCAAGATTTAGATTAATAATGAACGCTGT 896
QY 636 TGCTAA-----ACCAACGAAATTTATGTTAGTTGTGCGATTCATGAC 677


```
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-309-682-3

Query Match          7.5%; Score 102.4; DB 4; Length 831;
Best Local Similarity 51.5%; Pred. No. 6.4e-17;
Matches 300; Conservative 0; Mismatches 261; Indels 21; Gaps 2;

QY 364 SCATTATTGATCGCTAAATAAATACAAACAAACCTATGTTGTCAGCAGATATTAT 423
Db 280 GC---CCACCGCTACAAACGAACCTGGTAAGAAGCTCATGCTGTTGTCAGCAGATACCTTC 336
QY 424 COTCCAGCAGCGATCAATCAATTAACAAACAAACCTATGTTGTCAGCAGATATTAT 423
Db 337 CGTGGGGTGCAGTAGATCAGCTAGCTGAATGGGGCGCAGCAGTAGATGTTCCAGTAGTA 396
QY 484 AGTGAAGAGATCAAGTAAAGCCACACAAATTTCTAATCAATCAATCAATCAATCAAT 543
Db 337 CGTGGGGTGCAGTAGATCAGCTAGCTGAATGGGGCGCAGCAGTAGATGTTCCAGTAGTA 396
QY 484 AGTGAAGAGATCAAGTAAAGCCACACAAATTTCTAATCAATCAATCAATCAATCAAT 543
Db 397 ACTGACCTGAAAGAGTGAATGATCCAGCAGCGCTGTTGATGGTATGAAGAGTGCCTGTG 456
QY 544 GAAGAACATTTGACTTTGTAATCATTTGATACAGCAGCTGATTAACATCAATCAATCAAT 603
Db 457 GCTGAAGTATCGATATCTCATGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 516
QY 604 TTGATGAACAGATTAAGAAAGTAAAGAAATTTGCTAA-----ACCA 645
Db 517 CTTATGGCTGAGTGGAAAGATTTGGTCTGTTATTCATCAACGTTGTTGCCAGAACCA 576
QY 646 AACGAATTTAGTTAGTTGCGATTCATGAGCGGTCAAGATGCTGCTCAATCTTCAGAA 705
Db 577 CATGAACCTTCTTGGCAGCTTGATGCTACAGAGTCAGAGTCAGAGTCAGAGTCAGAG 636
QY 706 TCTTTTACAGATCAACTGATGCTACAGAGTGTACCTTAACATAAATAGATGGTGATACA 765
Db 637 GAATTTGGAATAATCACACCTTTACGGGGAATTTTGTGACTAAGATTTGATGAACCTCT 696
QY 766 CGTGTGTCAGCCTTTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
Db 697 CGAGGAGGTGTGTTCTAGCCATTCGTGAAGAACTCAATATTCCTGTAATAATGATTGGT 756
QY 826 ATGAGTGAAGAGTTAGATGTTGTTAGAGCTATTCAGATTCCTCTGAA 867
Db 757 TTTGTTGAAAAATCGATGATATTTGAGAGTTTAACTCAGAA 798

RESULT 14
US-09-309-682-3
; Sequence 3, Application US/09309682
; Patent No. 6214348
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6214348el Ftsy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/007,484
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T.
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GML0081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
```

```
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-309-682-3

Query Match          7.5%; Score 102.4; DB 4; Length 831;
Best Local Similarity 51.5%; Pred. No. 6.4e-17;
Matches 300; Conservative 0; Mismatches 261; Indels 21; Gaps 2;

QY 304 ACTGTTGTTATGATGTTGTTTACAAGTGTCTGGTAAACAAACAACTCCAGTAAATTA 363
Db 220 ACAGTATGCTCTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 279
QY 364 GCATTATTGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 423
Db 280 GC---CCACCGCTACAAACGAACCTGGTAAAGAGTCTATGCTGTTGCCAGCAGTACCTTC 336
QY 424 COTCCAGCAGCGATTAATCAATTAACAAACAAACCTGTTGTTGTTGTTGTTGTTGTTGTTG 483
Db 337 CGTGGGGTGCAGTAGATCAGCTAGCTGAATGGGGCGCAGCAGTAGATGTTCCAGTAGTA 396
QY 484 AGTGAAGAGATCAAGTAAAGCCACACAAATTTCTAATCAATCAATCAATCAATCAAT 543
Db 397 ACTGACCTGAAAGAGTGAATGATCCAGCAGCGCTGTTGATGGTATGAAGAGTGCCTGTG 456
QY 544 GAAGAACATTTGACTTTGTAATCATTTGATACAGCAGCTGATTAACATCAATCAATCAAT 603
Db 457 GCTGAAGTATCGATATCTCATGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 516
QY 604 TTGATGAACAGATTAAGAAAGTAAAGAAATTTGCTAA-----ACCA 645
Db 517 CTTATGGCTGAGTGGAAAGATTTGGTCTGTTATTCATCAACGCTGTTGCCAGAACCA 576
QY 646 AACGAATTTAGTTAGTTGCGATTCATGAGCGGTCAAGATGCTGCTCAATCTTCAGAA 705
Db 577 CATGAACCTTCTTGGCAGCTTGATGCTACAGAGTCAGAGTCAGAGTCAGAGTCAGAG 636
QY 706 TCTTTTACAGATCAACTGATGCTACAGAGTGTACCTTAACATAAATAGATGGTGATACA 765
Db 637 GAATTTGGAATAATCACACCTTTACGGGGAATTTTGTGACTAAGATTTGATGAACCTCT 696
QY 766 CGTGTGTCAGCCTTTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
Db 697 CGAGGAGGTGTGTTCTAGCCATTCGTGAAGAACTCAATATTCCTGTAATAATGATTGGT 756
QY 826 ATGAGTGAAGAGTTAGATGTTGTTAGAGCTATTCAGATTCCTCTGAA 867
Db 757 TTTGTTGAAAAATCGATGATATTTGAGAGTTTAACTCAGAA 798
```

```
RESULT 15
US-09-963-1
; Sequence 1, Application US/08986963
; Patent No. 5958730
; GENERAL INFORMATION:
; APPLICANT: Rostek Jr., Paul R.
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: Ftsy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,963
; FILING DATE: December 8, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1278 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1278
; US-08-986-963-1

Query Match          7.5%; Score 102.4; DB 2; Length 1278;
Best Local Similarity 51.5%; Pred. No. 7.5e-17;
Matches 300; Conservative 0; Mismatches 261; Indels 21; Gaps 2;

QY 304 ACCTGTTGTTATGATGGTTGTTACAGGTGGTGTGTTAAACAAACAACTGCAGGTAAATTA 363
Db 664 ACAGTTATGCTCTTGTGGTGTGATGGTGTGGGAAACAACTCTCTATCGGAAACTA 723

QY 364 GCATTTATTCGCTAAATAATACACAAACAACTATGTTAGTTCAGCAGATATTTAT 423
Db 724 GC---CCACCGGTACAAACAAAGCTGTAAGAGGTCATGCTGGTTGCAGCAGATACCTTC 780

QY 424 CGTCCAGCAGCGATTAATACAAACAGTAGGAAACAAATTTGATATTCTCGTATAC 483
Db 781 CGTGGGGTGGAGTAGTACGTACGTAGTGTGGGCGCGAGTAGATGTTCCAGTAGTA 840

QY 484 AGTGAAGGAGATCAAGTAAAGCCCAACAAATTTGAATGATTAATGATTAACATGCTAAA 543
Db 841 ACTGGACCTGAAAAGCTGATCCAGCCAGCGGTGCTGTTGATGGTATGGAACTGCCGTG 900

QY 544 GAGACACATTTAGACTTTGTAATCATTTGATACAGCAGGTGCGATTACACATCGATGAAGCA 603
Db 901 GCTGAAGGTATCGATTTCTCATGATTGATCTGCTGGTCTGCTGCAAAATAAGGATAAC 960

QY 604 TTGATGAACGAATTAAGAAAGTAAAGAAATTTGCTAA-----ACCA 645
Db 961 CTTATGGCTGAGTTGGAAAGATTTGGTCTGTAATTCACACGTGTTGTCCAGAGCAACA 1020

QY 646 AAGCAATATATTAGTTGTCGATTCATTAAGCAGGGTCAAGATGCTGTCATGTCAGAA 705
Db 1021 CATGAACCTTCTGGCATTGATGCATCAACAGGTCAAAATGCTGCTAGTACAGGCCAAA 1080

QY 706 TCTTTTCAGCATCACTTGATGTCACAGGTGTTACCTTAACATAATTAGATGGTGATACA 765
Db 1081 GAATTTTCGAAATCACACACCTTTACCGGAATTTGTTTGAATTAAGATTTGGAAGTCT 1140

QY 766 CGTGGTGGTGCAGCTTATCTATCTGTTCCGTGCACACAAAAACCAATTAATTTGTTGGT 825
Db 1141 CGAGGAGGTGGTCTCTAGCCATCTCGTGAAGAACTCAATATCTCTGTAATAATGATTTGT 1200

QY 826 ATGAGTGAAGAAGTTAGATGGTTAGAGCTATTCCATCCTGAA 867
Db 1201 TTTGGTGAAGAAATCGATGATATTGGAGAGTTTAACTCAGAA 1242
```

Search completed: February 25, 2003, 00:20:53
Job time : 106 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 23:49:06 ; Search time 75 Seconds
(without alignments)
10244.303 Million cell updates/sec

Title: US-09-943-108A-1
Perfect score: 1368
Sequence: 1 atggcatttgagccttatc.....gtatgaatttaacggttttaa 1368

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications,NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1356.8	99.2	1368	10	US-09-815-242-8968
2	1344	98.2	1368	10	US-09-815-242-8030
3	1341	98.0	1365	10	US-09-815-242-4365
4	703.4	51.4	721	10	US-09-815-242-1949
5	568.6	41.6	1416	10	US-09-815-242-3890
6	568.6	41.6	1434	10	US-09-815-242-6764
7	565	41.3	6729	10	US-09-070-927A-37
8	516.2	37.7	521	10	US-09-815-242-1985
9	516.2	37.7	521	10	US-09-815-242-2013
10	496	36.3	1395	10	US-09-815-242-9427
11	496	36.3	1569	10	US-09-814-041A-1
12	496	36.3	1572	10	US-09-815-242-9156
13	486.4	35.6	1136	10	US-09-974-300-1847
14	463.2	33.9	468	10	US-09-815-242-3070
15	443.4	32.4	1389	10	US-09-815-242-6868
16	418.2	30.6	640681	10	US-09-790-988-1
17	370.6	27.1	1362	10	US-09-815-242-6166
18	360	26.3	1506	10	US-09-815-242-9391
19	328.4	24.0	330	10	US-09-815-242-2247

20	280	20.5	1374	10	US-09-815-242-7850
21	263.6	19.3	292	10	US-09-815-242-3369
22	259	18.9	1641	9	US-09-738-626-2255
23	254	18.6	1347	10	US-09-815-242-7465
24	160.4	11.7	1440	9	US-09-938-842A-877
25	145.6	10.6	1251	10	US-09-815-242-8028
26	143.2	10.5	948	10	US-09-815-242-4244
27	133	9.7	939	10	US-09-815-242-3937
28	133	9.7	1329	10	US-09-815-242-6807
29	133	9.7	15363	10	US-09-070-927A-279
30	102.4	7.5	831	10	US-09-827-663-3
31	102.4	7.5	1278	10	US-09-815-242-9165
32	100.8	7.4	1293	10	US-09-827-663-1
33	97.6	7.1	1290	10	US-09-815-242-9491
34	95.6	7.0	643	10	US-09-815-242-2506
35	82.6	6.0	1245	10	US-09-815-242-6997
36	64	4.7	1494	10	US-09-815-242-6281
37	63.8	4.7	405	10	US-09-960-352-13494
38	60.4	4.4	218	10	US-09-815-242-534
39	58	4.2	29793	10	US-09-973-451-38
40	55.6	4.1	610	9	US-09-736-457-1094
41	55.6	4.1	610	9	US-09-902-941-1094
42	55.6	4.1	610	9	US-09-849-626-1094
43	54.4	4.0	1530	9	US-09-738-626-2259
44	52.2	3.8	774	10	US-09-922-261-187
45	52.2	3.8	819	10	US-09-922-261-185

ALIGNMENTS

RESULT 1

US-09-815-242-8968
; Sequence 8968, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8968
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8968

Sequence 7850, Ap
Sequence 3369, Ap
Sequence 2255, Ap
Sequence 7465, Ap
Sequence 877, Ap
Sequence 8028, Ap
Sequence 4244, Ap
Sequence 3937, Ap
Sequence 6807, Ap
Sequence 279, Ap
Sequence 3, Appli
Sequence 9165, Ap
Sequence 1, Appli
Sequence 9491, Ap
Sequence 2506, Ap
Sequence 6997, Ap
Sequence 6281, Ap
Sequence 13494, A
Sequence 534, App
Sequence 38, Appl
Sequence 1094, Ap
Sequence 1094, Ap
Sequence 2259, Ap
Sequence 187, App
Sequence 185, App

Query Match				99.2%; Score 1356.8; DB 10; Length 1368;			
Best Local Similarity				99.5%; Pred. No. 3.2e-287;			
Matches 1361; Conservative				0; Mismatches 7; Indels 0; Gaps 0;			
QY	1	ATGCGATTGTAAGGCTTATCAGAACGCTGCAAGCGAGATGCAAAAATATGCGTGTAAAG	60				
Db	1	ATGCGATTGTAAGGCTTATCAGAACGCTTCAAGCGAGATGCAAAAATATGCGTGTAAAG	60				
QY	61	GSTAACTTACGAGCTGATTAAGATAATGATCGCTGAGTAAGATTAAGGTTATTT	120				
Db	61	GSTAACTTACGAGCTGATTAAGATAATGATCGCTGAGTAAGATTAAGGTTATTT	120				
QY	121	GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACCGCA	180				
Db	121	GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGAACCGCA	180				
QY	181	TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGACAGTATTAATAAGTTCAA	240				
Db	181	TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGACAGTATTAATAAGTTCAA	240				
QY	241	GATGAATTAACCAAGCTGATGGGCGGAGAAATACATCGATTATATGTCAAATAAACCA	300				
Db	241	GATGAATTAACCAAGCTGATGGGCGGAGAAATACATCGATTATATGTCAAATAAACCA	300				
QY	301	CTTACTGTTTATGATGCTGTTGTTTACAAAGTGGTGTAAACCAACAACTGCAGGTAA	360				
Db	301	CTTACTGTTTATGATGCTGTTGTTTACAAAGTGGTGTAAACCAACAACTGCAGGTAA	360				
QY	361	TTAGCATTTATGATCGTAAATAATACACAAAACCTATGTTAGTTGCAGCAGATAT	420				
Db	361	TTAGCATTTATGATCGTAAATAATACACAAAACCTATGTTAGTTGCAGCAGATAT	420				
QY	421	TATCGTCCAGCAGGATAAATCAATTAACAACAGTAGGAAACAAATGATATTCGTGA	480				
Db	421	TATCGTCCAGCAGGATAAATCAATTAACAACAGTAGGAAACAAATGATATTCGTGA	480				
QY	481	TACAGTGAAGGAGATCAAGTAAAGCAACAAATTTGAACCTAATGATTAACAAACATG	540				
Db	481	TACAGTGAAGGAGATCAAGTAAAGCAACAAATTTGAACCTAATGATTAACAAACATG	540				
QY	541	AAAGAGACATTTAGATTTGTTATCATTTGATACAGAGTGCATACACATCGATGAA	600				
Db	541	AAAGAGACATTTAGATTTGTTATCATTTGATACAGAGTGCATACACATCGATGAA	600				
QY	601	GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAAACCAAGAAATTTATGTTA	660				
Db	601	GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAAACCAAGAAATTTATGTTA	660				
QY	661	GTTGTCGATTCATGACGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTGACGATCAA	720				
Db	661	GTTGTCGATTCATGACGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTGACGATCAA	720				
QY	721	CTTGATCTCACAGGCTTACCTTAACATAATAGATGCTGATACACGTTGGTGGTGCAGCT	780				
Db	721	CTTGATCTCACAGGCTTACCTTAACATAATAGATGCTGATACACGTTGGTGGTGCAGCT	780				
QY	781	TTATCTATTCGTTGCTGACACAAAACCAATTAATTTGTTGATGAGTGAAGTTA	840				
Db	781	TTATCTATTCGTTGCTGACACAAAACCAATTAATTTGTTGATGAGTGAAGTTA	840				
QY	841	GATGGTTTAGAGCTATTCCTCTGAACGATGCGCATCAGTATTTTAGTATGGGTGAT	900				
Db	841	GATGGTTTAGAGCTATTCCTCTGAACGATGCGCATCAGTATTTTAGTATGGGTGAT	900				
QY	901	GTGTTAAGTTTAAATGAAGACCGACACAGATGTGGATCAAGAAAGCAAGATTTA	960				
Db	901	GTGTTAAGTTTAAATGAAGACCGACACAGATGTGGATCAAGAAAGCAAGATTTA	960				
QY	961	GAGAAAAGATGGGTGAGTCATCGTTTACTTTTAGATGATTTTTTAAAGCAACTGTGATC	1020				
Db	961	GAGAAAAGATGGGTGAGTCATCGTTTACTTTTAGATGATTTTTTAAAGCAACTGTGATC	1020				

QY	1021	GTGAAAAATCTAGGACCACCTGGATGATATATATGAAAAATGATTCAGGTATGATAATAATG	1080				
Db	1021	GTGAAAAATCTAGGACCACCTGGATGATATATATGAAAAATGATTCAGGTATGATAATAATG	1080				
QY	1081	AAAGGCTGATAGTAAGCTTAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATTATC	1140				
Db	1081	AAAGGCTGATAGTAAGCTTAATATGAGTGAAGCAAAATTTGATCATATTAAGCGATTATC	1140				
QY	1141	CATCAATCAGCCCGCTGAAAGAAACAATCCAGACACATTTGAATGTATCACGTAAAAAG	1200				
Db	1141	CATCAATCAGCCCGCTGAAAGAAACAATCCAGACACATTTGAATGTATCACGTAAAAAG	1200				
QY	1201	CGTATTCTAAAGGCTCTGTCGTTTCATTTACAGAGTCAATCGTTTGCATCAACAATTT	1260				
Db	1201	CGTATTCTAAAGGCTCTGTCGTTTCATTTACAGAGTCAATCGTTTGCATCAACAATTT	1260				
QY	1261	AAGCATATGAAGAAATGATGACACATTTCACTGTGGCGGTAAAGGTAAAAAGGTAAA	1320				
Db	1261	AAGCATATGAAGAAATGATGACACATTTCACTGTGGCGGTAAAGGTAAAAAGGTAAA	1320				
QY	1321	CGCAATCAAAATGCAAAATATGTTAAAAAGTATGAATTTACCGTTTTAA	1368				
Db	1321	CGCAATCAAAATGCAAAATATGTTAAAAAGTATGAATTTACCGTTTTAA	1368				

RESULT 2

US-09-815-242-8030
; Sequence 8030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8030
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8030

Query Match 98.2%; Score 1344; DB 10; Length 1368;
Best Local Similarity 98.9%; Pred. No. 2e-284;
Matches 1353; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGCGATTGAGGCTTATCAGAACGCTGCAAGCGATGCAAAAATATGCGTGTAAAG 60

|||||
Db 1 ATGGCAATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60
QY 61 GGTAACTTACTGAAGCTGATATGAAGATATGATGCGTGAAGTATGATGCGTGAATTT 120
Db 61 GGTAACTTACTGAAGCTGATATGAAGATATGATGCGTGAAGTATGATGCGTGAATTT 120
QY 121 GAGGTGACGTAAAGCTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGGCA 180
Db 121 GAGGTGACGTAAAGCTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGGCA 180
QY 181 TTAGTTCGCGATTAATCAATCATTAAACAGGCGCAACAGTTATTAATAGTTCAA 240
Db 181 TTAGTTCGCGATTAATCAATCATTAAACAGGCGCAACAGTTATTAATAGTTCAA 240
QY 241 GATGAATTAACAGGTTGATGCGTGAAGAAATATCATCGATTAATATGTCAAATAACCA 300
Db 241 GATGAATTAACAGGTTGATGCGTGAAGAAATATCATCGATTAATATGTCAAATAACCA 300
QY 301 CTTACTGTGTATGATGTTGTTTACAAGGTGCTGTTAAACAACTTCAGGTA 360
Db 301 CTTACTGTGTATGATGTTGTTTACAAGGTGCTGTTAAACAACTTCAGGTA 360
QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTTATGTTGCGACGATATT 420
Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTTATGTTGCGACGATATT 420
QY 421 TATCTCCAGCAGGATTAATCAATTAACACAGTATGAGGAAACAAATGATTCCTGTA 480
Db 421 TATCTCCAGCAGGATTAATCAATTAACACAGTATGAGGAAACAAATGATTCCTGTA 480
QY 481 TACGTGAAGGATCAAGTAAAGCCACAAATTTGTAATCAATCAATCAATCAATCAATGCT 540
Db 481 TACGTGAAGGATCAAGTAAAGCCACAAATTTGTAATCAATCAATCAATCAATCAATGCT 540
QY 541 AAAGAAGCAATTTAGACTTTTGAATTCATGATACAGCAGGTCGATACACATCATGATGAA 600
Db 541 AAAGAAGCAATTTAGACTTTTGAATTCATGATACAGCAGGTCGATACACATCATGATGAA 600
QY 601 GCATTGATGAATTAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660
Db 601 GCATTGATGAATTAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660
QY 661 GTTGTGCAATCAATGAGGTCAGATGCTGCTCAATGTTTCAGAAATCTTTTACAGATCAA 720
Db 661 GTTGTGCAATCAATGAGGTCAGATGCTGCTCAATGTTTCAGAAATCTTTTACAGATCAA 720
QY 721 CTTGATGTCACAGGTTTACCTTAATAAATAGATGTTGATACAGTGGTGGTGCAGCT 780
Db 721 CTTGATGTCACAGGTTTACCTTAATAAATAGATGTTGATACAGTGGTGGTGCAGCT 780
QY 781 TTATCTATGCTGCGTGACCAAAACCAATTAATTTTGGTATGAGTGAAGATTA 840
Db 781 TTATCTATGCTGCGTGACCAAAACCAATTAATTTTGGTATGAGTGAAGATTA 840
QY 841 GATGTTTATGAGCTATTCATCCTGAAGTATGCTATCAATGTTTATGATGAGTGGTGAT 900
Db 841 GATGTTTATGAGCTATTCATCCTGAAGTATGCTATCAATGTTTATGATGAGTGGTGAT 900
QY 901 GTGTTAAGTTTAAATGAAGGCGCAACAGATGTTGATCAAGAAACCAAAAGATTTA 960
Db 901 GTGTTAAGTTTAAATGAAGGCGCAACAGATGTTGATCAAGAAACCAAAAGATTTA 960
QY 961 GAGAAAAGATGCGTGAAGTATGCTTTTATGATGATTTTATGATGATTTTATGATGATGATG 1020
Db 961 GAGAAAAGATGCGTGAAGTATGCTTTTATGATGATTTTATGATGATTTTATGATGATGATG 1020
QY 1021 GTGAAAATCTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Db 1021 GTGAAAATCTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1081 AAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140

Db 1081 AAAGTCTAGATAGCTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
QY 1141 CAGTCAATGACGCGCTGAAGAAACAAATCCAGACATTTGAATGATCAATGATGATGATGATG 1200
Db 1141 CAGTCAATGACGCGCTGAAGAAACAAATCCAGACATTTGAATGATCAATGATGATGATGATG 1200
QY 1201 CGTATTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1201 CGTATTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1261 AACGATATGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
Db 1261 AACGATATGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
QY 1321 CGCATCAATGCAAAATATGTTTAAAGGTATGATGATGATGATGATGATGATGATGATGATGATG 1368
Db 1321 CGCATCAATGCAAAATATGTTTAAAGGTATGATGATGATGATGATGATGATGATGATGATGATG 1368

RESULT 3
US-09-815-242-4365
; Sequence 4365, Application us/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Ohlsep, Robert
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4365
; LENGTH: 1365
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4365

Query Match 98.0%; Score 1341; DB 10; Length 1365;
Best Local Similarity 98.9%; Pred. No. 8.9e-284;
Matches 1350; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGCGATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60
Db 1 ATGCGATTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60
QY 61 GGTAACTTACTGAAGCTGATATGAAGATATGATGCGTGAAGTATGATGCGTGAATTT 120
Db 61 GGTAACTTACTGAAGCTGATATGAAGATATGATGCGTGAAGTATGATGCGTGAATTT 120
QY 121 GAGGTGACGTAAAGCTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGGCA 180
Db 121 GAGGTGACGTAAAGCTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGACGGCA 180
QY 181 TTAGTTCGCGATTAATCAATCATTAAACAGGCGCAACAGTTATTAATAGTTCAA 240
Db 181 TTAGTTCGCGATTAATCAATCATTAAACAGGCGCAACAGTTATTAATAGTTCAA 240
QY 241 GATGAATTAACAGGTTGATGCGTGAAGAAATATCATCGATTAATATGTCAAATAACCA 300
Db 241 GATGAATTAACAGGTTGATGCGTGAAGAAATATCATCGATTAATATGTCAAATAACCA 300
QY 301 CTTACTGTGTATGATGTTGTTTACAAGGTGCTGTTAAACAACTTCAGGTA 360
Db 301 CTTACTGTGTATGATGTTGTTTACAAGGTGCTGTTAAACAACTTCAGGTA 360
QY 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTTATGTTGCGACGATATT 420
Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTTATGTTGCGACGATATT 420
QY 421 TATCTCCAGCAGGATTAATCAATTAACACAGTATGAGGAAACAAATGATTCCTGTA 480
Db 421 TATCTCCAGCAGGATTAATCAATTAACACAGTATGAGGAAACAAATGATTCCTGTA 480
QY 481 TACGTGAAGGATCAAGTAAAGCCACAAATTTGTAATCAATCAATCAATCAATGCT 540
Db 481 TACGTGAAGGATCAAGTAAAGCCACAAATTTGTAATCAATCAATCAATCAATGCT 540
QY 541 AAAGAAGCAATTTAGACTTTTGAATTCATGATACAGCAGGTCGATACACATCATGATGAA 600
Db 541 AAAGAAGCAATTTAGACTTTTGAATTCATGATACAGCAGGTCGATACACATCATGATGAA 600
QY 601 GCATTGATGAATTAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660
Db 601 GCATTGATGAATTAATTAAGAAAGTAAAGAAATTTGCTTAACCAACAAATTAATGTTA 660
QY 661 GTTGTGCAATCAATGAGGTCAGATGCTGCTCAATGTTTCAGAAATCTTTTACAGATCAA 720
Db 661 GTTGTGCAATCAATGAGGTCAGATGCTGCTCAATGTTTCAGAAATCTTTTACAGATCAA 720
QY 721 CTTGATGTCACAGGTTTACCTTAATAAATAGATGTTGATACAGTGGTGGTGCAGCT 780
Db 721 CTTGATGTCACAGGTTTACCTTAATAAATAGATGTTGATACAGTGGTGGTGCAGCT 780
QY 781 TTATCTATGCTGCGTGACCAAAACCAATTAATTTTGGTATGAGTGAAGATTA 840
Db 781 TTATCTATGCTGCGTGACCAAAACCAATTAATTTTGGTATGAGTGAAGATTA 840
QY 841 GATGTTTATGAGCTATTCATCCTGAAGTATGCTATCAATGTTTATGATGAGTGGTGAT 900
Db 841 GATGTTTATGAGCTATTCATCCTGAAGTATGCTATCAATGTTTATGATGAGTGGTGAT 900
QY 901 GTGTTAAGTTTAAATGAAGGCGCAACAGATGTTGATCAAGAAACCAAAAGATTTA 960
Db 901 GTGTTAAGTTTAAATGAAGGCGCAACAGATGTTGATCAAGAAACCAAAAGATTTA 960
QY 961 GAGAAAAGATGCGTGAAGTATGCTTTTATGATGATTTTATGATGATTTTATGATGATGATG 1020
Db 961 GAGAAAAGATGCGTGAAGTATGCTTTTATGATGATTTTATGATGATTTTATGATGATGATG 1020
QY 1021 GTGAAAATCTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Db 1021 GTGAAAATCTAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1081 AAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140

Db 121 GAGGCTGACGTAAACCTTTAAAGTGGTAAAGATTTTATTAACACAGATATCAGAACGCGCA 180
QY 181 TTAGGTTCCGATGTAATCAATCAATTAACACAGGCGCAACAGTTTATTAATAATAGTTCAA 240
Db 181 TTAGGTTCCGATGTAATGCAATCAATTAACACAGGCGCAACAGTTTATTAATAATAGTTCAA 240
QY 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATGTCTCAAAATAACCA 300
Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATATATCGATTATATGTCTCAAAATAACCA 300
QY 301 CCTACTGTTGTTATGATGGTGGTTTACAGGTCGTGGTAAACACACACTCGAGGTAAA 360
Db 301 CCTACTGTTGTTATGATGGTGGTTTACAGGTCGTGGTAAACACACACTCGAGGTAAA 360
QY 361 TTACGATTATGATCGTTAAATAATACACAAAAACCTTATCTAGTTGGCAGCAGATATT 420
Db 361 TTACGATTATGATCGTTAAATAATACACAAAAACCTTATCTAGTTGGCAGCAGATATT 420
QY 421 TATCGTCCAGCAGGATAAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTGTA 480
Db 421 TATCGTCCAGCAGGATAAATCAATTACAAACAGTAGGGAACAAATTTGATATTCCTGTA 480
QY 481 TACAGTGAAGGAGTCAAGTAAAGCCACACAAATTTGTAAGTATGCAATTAACAGATGCT 540
Db 481 TACAGTGAAGGAGTCAAGTAAAGCCACACAAATTTGTAAGTATGCAATTAACAGATGCT 540
QY 541 AAAGAGAACATTTAGACTTTTGAATCATTTGATACAGCAGGTCGATTACATCGATGAA 600
Db 541 AAAGAGAACATTTAGACTTTTGAATCATTTGATACAGCAGGTCGATTACATCGATGAA 600
QY 601 GCATTGATGAACGAATTAAGAGAGTAAAGAAATTTGCTTAAACCAAGCAAAATTTGTTA 660
Db 601 GCATTGATGAATTAAGAGAGTAAAGAAATTTGCTTAAACCAAGCAAAATTTGTTA 660
QY 661 GTTGTGATTCATGACGGGTCAAGATGCTGCAATGTTGCGAAGATCTTTTGACGATCAA 720
Db 661 GTTGTGATTCATGACGGGTCAAGATGCTGCAATGTTGCGAAGATCTTTTGACGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAACTAAATTAGATGGTGATACAGTGGTGGGAGCT 780
Db 721 CTTGATGTCACAGGTGTTACCTTAACTAAATTAGATGGTGATACAGTGGTGGGAGCT 780
QY 781 TTATCTATTGTTGCGGTGACACAAAACCAATTAATTTGTTGTPANGAGTGAAGTTA 840
Db 781 TTATCTATTGTTGCGGTGACACAAAACCAATTAATTTGTTGTPANGAGTGAAGTTA 840
QY 841 GATGGTTAGAGCTATTCATCCTGCAACGATATGGCATCAGTATTTTTRAGTATGGGTAT 900
Db 841 GATGGTTAGAGCTATTCATCCTGCAACGATATGGCATCAGTATTTTTRAGTATGGGTAT 900
QY 901 GTGTTAAGTTTAAATGAAAAGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960
Db 901 GTGTTAAGTTTAAATGAAAAGCGCAACAGATGTGGATCAAGAAAAGCAAAAGATTTA 960
QY 961 GAGAAAAGATGCGGTGAGTCATCGTTTACTTTTAGATGATTTTTTTTGAACAACATTTGATC 1020
Db 961 GAGAAAAGATGCGGTGAGTCATCGTTTACTTTTAGATGATTTTTTTTGAACAACATTTGATC 1020
QY 1021 GTGAAAATCTAGGACCACCTGGATGATATTTATGAAAATGATTCAGAGTATGAATAAATG 1080
Db 1021 GTGAAAATCTAGGACCACCTGGATGATATTTATGAAAATGATTCAGAGTATGAATAAATG 1080
QY 1081 AAAGGGCTAGATAGCTTAATGATGAGTGAAGCAAAATTCATATTAAGCGGATATTC 1140
Db 1081 AAAGGGCTAGATAGCTTAATGATGAGTGAAGCAAAATTCATATTAAGCGGATATTC 1140
QY 1141 CAGTCAATGACGCGGCTGAAGAAAACCAATCCAGACACATTTGAATGCTATCAGTAAAAAG 1200
Db 1141 CAGTCAATGACGCGGCTGAAGAAAACCAATCCAGACACATTTGAATGCTATCAGTAAAAAG 1200
QY 1201 CGTATTGCTAAAGGTCGTGGTTCATTAACAAGAGTCAATCGTTTGATGAACAATTT 1260
Db 1201 CGTATTGCTAAAGGTCGTGGTTCATTAACAAGAGTCAATCGTTTGATGAACAATTT 1260

QY 1261 AACGATATGAAGAAATGATGAACAATTCACGTGGCGGTAAAGGTAAAAAGGTAAA 1320
Db 1261 AACGATATGAAGAAATGATGAACAATTCACGTGGCGGTAAAGGTAAAAAGGTAAA 1320
QY 1321 CGCAATCAAAATCAAAATATGTTTAAAGGTATGAATTTACCGTTT 1365
Db 1321 CGCAATCAAAATCAAAATATGTTTAAAGGTATGAATTTACCGTTT 1365

RESULT 4
US-09-815-242-1949/c
; Sequence 1949, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1949
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-1949

Query Match 51.4%; Score 703.4; DB 10; Length 721;
Best Local Similarity 99.2%; Pred. No. 1.1e-144;
Matches 707; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGCATTTGAAGGCTTATCAGAACGCGCTGCAAGCGACGATCAAAAAATCGTGTGAAG 60
Db 713 ATGCATTTGAAGGCTTATCAGAACGCGCTTGCAGCGACGATCAAAAAATCGTGTGAAG 654
QY 61 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTCAAGTAAAGTTAGCGTTATT 120
Db 653 GGTAACTTACTGAAGCTGATATAAGATAATGATCGGTCAAGTAAAGTTAGCGTTATT 594
QY 121 GAGGCTCAGCTAACTTTTAAAGTGTAAAGAAATTTTAAACACAGTATCAGAACGCGCA 180
Db 593 GAGGCTCAGCTAACTTTTAAAGTGTAAAGAAATTTTAAACACAGTATCAGAACGCGCA 534
QY 181 TTAGGTTCCGATGTAATGCAATCAATTAACACAGGCAACAAAGTTTATTAATAATAGTTCAA 240
Db 533 TTAGGTTCCGATGTAATGCAATCAATTAACACAGGCAACAAAGTTTATTAATAATAGTTCAA 474
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTATATGTCAAATAACCA 300
Db 473 GATGAATTAACGAAGTTGATGGTGGAGAAAATACATCGATTATATGTCAAATAACCA 414

QY 301 CCTACTGTTCTTATGATGGTTGGTTTACAAAGGTGCTGGTAAACACAACTCCAGGTAAA 360
DB 413 CCTACTGTTCTTATGATGGTTGGTTTACAAAGGTGCTGGTAAACACAACTCCAGGTAAA 354
QY 361 TTAGCATTTATGATGGTAAAGAAATACAAACAAACAACTAGTTAGTTCCAGCAGATATT 420
DB 353 TTAGCATTTATGATGGTAAAGAAATACAAACAAACAACTAGTTAGTTCCAGCAGATATT 294
QY 421 TATGTCGAGCAGGATTAATCAATTAACAAACAGTAGGGAACAAATTTGATATCTCTGTA 480
DB 293 TATGTCGAGCAGGATTAATCAATTAACAAACAGTAGGGAACAAATTTGATATCTCTGTA 234
QY 481 TACAGTAGAGAGATCAAGTAAGCCACACAAATTTGTAATCAATTAACAAATGCT 540
DB 233 TACAGTAGAGAGATCAAGTAAGCCACACAAATTTGTAATCAATTAACAAATGCT 174
QY 541 AAAGAAAGCAATTTAGATCTTTGTAATCAATTTGATACAGAGCTGATACACATCGATGAA 600
DB 173 AAAGAAAGCAATTTAGATCTTTGTAATCAATTTGATACAGAGCTGATACACATCGATGAA 114
QY 601 GCATTTGATGACGAAATTAAGAAAGTAAAGAAATTTGCTAAACCAACGAAATTTATGTTA 660
DB 113 GCATTTGATGACGAAATTAAGAAAGTAAAGAAATTTGCTAAACCAACGAAATTTATGTTA 54
QY 661 GTTGTGATTCATCAAGCGGTGCAAGATGCTGTCAATGTTSCAATCTTTTGA 713
DB 53 GTTGTGATTCATCAAGCGGTGCAAGATGCTGTCAATGTTSCAATCTTTTGA 1

RESULT 5

US-09-815-242-3890
; Sequence 3890, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; CURRENT APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3890
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3890

Query Match 41.6%; Score 568.6; DB 10; Length 1416;
Best Local Similarity 64.8%; Pred. No. 3.4e-115;
Matches 844; Conservative 0; Mismatches 459; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGCTTTATCAGAACCCCTGCAAGCAGATGCAAAAAAATGCGTGGTAA 60
DB 1 ATGGCTTTTGAAGCTTTTAAACAAACCCCTTACACAGCAGATGATTAATAATCCGCTGTA 60
QY 61 GCTAAACCTTACTGAAGCTGATATAAGAGATAATGAGCGTGAAGTAAAGATAGCGTTATTT 120
DB 61 GGAAGAGTTCCGGAAGCCGACGTTAAAGAAATGATGCGAGAAATCCGTTTGGCTTTATTA 120
QY 121 GAGGTGACGTAAATTTAAAGTGTAAAAAGAAATTTATTAACACAGTATCATCAGACGCGCA 180
DB 121 GAAGCCGACGTTAATTTACAAAGTGTCAAGATTTTCAAAAACCCGCTCAGAGACGCGCA 180
QY 181 TTAGTTCGCGATGTAATCAATCAITTAACACAGGCGCAACAAGTTATTAATAATGTTCAA 240
DB 181 GTAGGAGTCGAAGATTTAGAAAGCTTTATCACCAGCCCAACAAATTTGTAATAATTTGAT 240
QY 241 GATGAATTAACGAAGTTGATGGGTGAGAAATACATCGATTAATATGTCAAATAAACCA 300
DB 241 GAGAATTAACGAAGCTTTAGGTTTCAAGAAACGTTGAACCTGAATTAATCTCCAAAAATC 300
QY 301 CCTACTGTTGTTATGATGGTTGGTTTACAAAGGTGCTGTTAAAAACAACTACAGGTAAA 360
DB 301 CCGACAGTGATTTATGATGACAGGCTTTACAAAGGCTGGTAAACAACTTTACTGTTAAA 360
QY 361 TTAGCATTTATGATGGCTTAAAGAAATACAAACAAACCTATGTTAGTTGCAGCAGATATT 420
DB 361 TTAGCAAAACACTTTAATGAAACACTGAAACGCTGCTCGCTTTTAAATCGCTGGTACCT 420
QY 421 TATCGTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATTTGATATCTCTGTA 480
DB 421 TATCGTCCAGCAGGATTTGATGATGAGGTTTGAAGTTTGAATCAATTAAGAGTTCCGTT 480
QY 481 TACAGTAGAGAGATCAAGTAAGCCACACAAATTTGTAATCAATTAAGAGTTCCGTT 540
DB 481 TTTGATATGGGACAGATGCTAATCCAGTGGAAATTTGCTCAAGGTTAGCATTAGCA 540
QY 541 AAAGAAAGCAATTTAGACTTTTGAATCATTGATACAGCAGGCTGATTAACATCAATGAA 600
DB 541 AAAGAAAGCAATTTAGATTTTAAATGATACGCGCGCTTTTACATTTACATTTGACGAA 600
QY 601 GCATTTGATGAAGCAATTAAGAAAGTAAAGAAATTTGCTAAACCAACGAAATTTGTTA 660
DB 601 GCTTTAATGGAGCAATTTGAACAAATTAAGAGTTGGCTAATCCCAATGAATCTGTTA 660
QY 661 GTTGTGATTCATGACGCGGTCAAGATGCTGTCATTTGTCAGAAATCTTTTACGATCAA 720
DB 661 GTTGTGATGCGATGACGCGGCAAGATGCTGTCACAGTTGCGAGATAGTTTAAATGAACAG 720
QY 721 CTGTGATGTCACAGGTGTACCTTAACATAATTAAGATGATGATACAGTGGTGCAGCT 780
DB 721 CTGTGAATTTACTGGGTTGTTTATCCAAATGGACGCGGCTGCTGGGGGGGCTGCG 780
QY 781 TTATCTATTTCGTTGTCACACAAACCAATTAATTTGTTGATGATGATGATGAAAGTTA 840
DB 781 CTGTCAATTTGCGGCAAGTAAACGCGGCTCCGATTAATTTGTCGTTCTGGTGAATAATTA 840
QY 841 GATGTTTATGAGCTATTCTCTGCAACGATGCGATCAGCTATTTTATGATGATGATGAT 900
DB 841 ACCGATTTAGAAATTTTCCATCCGATCGTATGTCGAGTCTGATCTCTGATGATGATGAT 900
QY 901 GTGTTAAGTTTAATTAAGAAAGCGCAAGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 ATGTTGACGCTAATTTGAAAGCGCAAGATTTACGATGATGATGATGATGATGATGATGAT 960
QY 961 GAGAAAAAGATGCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 GCTCAAAAAATGAAGAAACAGCTTTTGGCTTAAAGATTTTCAATGATGATGATGATGAT 1020
QY 1021 GTCAAAAAATCTAGGACCACTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
DB 1021 GTTATGGCATGGGACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

Db 4982 AAGAAATCCAAAAAATGATGACAAATGTCCAAAGGGGATA 5024

RESULT 8

US-09-815-242-1985/c

; Sequence 1985, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1985

; LENGTH: 521

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-1985

Query Match 37.7%; Score 516.2; DB 10; Length 521;

Best Local Similarity 99.4%; Pred. No. 7.1e-104;

Matches 518; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 59 AGGTAACCTTACTGAAGCTGATTAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 118

Db 521 AGGTAACCTTACTGAAGCTGATTAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAC 462

QY 119 TTGAGGCTGACGTAACCTTAAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 178

Db 461 TTGAGGCTGACGTAACCTTAAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAC 402

QY 179 CATTAGTTCGCGATGTAATGCAATTAACACAGGCGCAACAAGTTATTAATAATAGTTTC 238

Db 401 CATTAGTTCGCGATGTAATGCAATTAACACAGGCGCAACAAGTTATTAATAATAGTTTC 342

QY 239 AAGTGAATTAACGAGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 298

Db 341 AAGTGAATTAACGAGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 282

QY 299 CACCTACTGTGTGTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 358

Db 261 CACCTACTGTGTGTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 222

QY 359 AATTAGCATTTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 418

Db 221 AATTAGCATTTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 162

QY 419 TTTATCGTCCAGCAGGATTAATCAATTTACAACTAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 478

Db 419 TTTATCGTCCAGCAGGATTAATCAATTTACAACTAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 478

Db 161 TTTATCGTCCAGCAGGATTAATCAATTTACAACTAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 102

QY 479 TATACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATTAATGCAATTAATAAATG 538

Db 101 TATACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATTAATGCAATTAATAAATG 42

QY 539 CTAAGAAGAACAATTTAGACTTTTGAATCATTTGATACAGCA 579

Db 41 CTAAGAAGAACAATTTAGACTTTTGAATCATTTGATACAGCA 1

RESULT 9

US-09-815-242-2013/c

; Sequence 2013, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2013

; LENGTH: 521

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-2013

Query Match 37.7%; Score 516.2; DB 10; Length 521;

Best Local Similarity 99.4%; Pred. No. 7.1e-104;

Matches 518; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 59 AGGTAACCTTACTGAAGCTGATTAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 118

Db 521 AGGTAACCTTACTGAAGCTGATTAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAC 462

QY 119 TTGAGGCTGACGTAACCTTAAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 178

Db 461 TTGAGGCTGACGTAACCTTAAAGTGTAAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 402

QY 179 CATTAGTTCGCGATGTAATGCAATTAACACAGGCGCAACAAGTTATTAATAATAGTTTC 238

Db 401 CATTAGTTCGCGATGTAATGCAATTAACACAGGCGCAACAAGTTATTAATAATAGTTTC 342

QY 239 AAGTGAATTAACGAGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 298

Db 341 AAGTGAATTAACGAGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 282

QY 299 CACCTACTGTGTGTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 358

Db 419 CACCTACTGTGTGTATGATGCGTGTGAGGTAAGATAATGATGCGTGAAGTAAGATTACCGTTAT 358

Db 281 CACCTACTGTTGTTATGCTGGTTGGTTTACAAAGCTGCTGTAACAAACACAACTGCAGSTA 222
QY 359 AATTAGCATATTTCATCGCTGCTAATAAATACAAACAAACCTATGTTAGTTCCACAGATA 418
Db 221 AATTAGCATATTTCATCGCTGCTAATAAATACAAACAAACCTATGTTAGTTCCACAGATA 162
QY 419 TTTATCGTCCAGCAGGATTAATTAATTAACAAACAGTAGGGAACAAATTAATGATTTCTG 478
Db 161 TTTATCGTCCAGCAGGATTAATTAATTAACAAACAGTAGGGAACAAATTAATGATTTCTG 102
QY 479 TATACAGTAAGGAGATCAAGTAAGCCACCAACAAATTAATGATTAATTAATGATTTCTG 538
Db 101 TATACAGTAAGGAGATCAAGTAAGCCACCAACAAATTAATGATTAATTAATGATTTCTG 42
QY 539 CTAAAGAGACATTTAGACTTTGTAATCAATGATACAGCA 579
Db 41 CTAAAGAGACATTTAGACTTTGTAATCAATGATACAGCA 1

RESULT 10
US-09-815-242-9427
; Sequence 9427, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1395)
US-09-815-242-9427

Query Match 36.3%; Score 496; DB 10; Length 1395;
Best Local Similarity 61.1%; Pred. No. 2.4e-99;
Matches 802; Conservative 0; Mismatches 510; Indels 0; Caps 0;

QY 1 ATGGCAATTGAGGCTTATCAGACGCTGCAAGCGACGTGCACAAATATGCGTGAAG 60
Db 1 ATGGCAATTGAGGCTTATCAGACGCTTGCAGACGCTTCTTTAAATATCTACGTAACAAA 60
QY 61 GGTAAATCTACTGAAGCTGATATAAGATAATGATGCGTGAAGTGAAGATTAAGCTTATTT 120
Db 61 GGTAAATCTACTGAATCTGATGTCCAAAGGCAACCAAGAAATTCGCTTGCCTTGCCTC 120

QY 121 GAGGCTGACGTAACACTTTAAAGTGGTAAAGAAATTTATTAACACATATCATCAAGCGCA 180
Db 121 GAGGCTGACGTAACACTTTAAAGTGGTAAAGAAATTTATTAACACATATCATCAAGCGTCA 180
QY 181 TTAGTCTCCGATGTAATGCAATCAATTAACACAGGCAACAACTTATTAATAATAGTTCAA 240
Db 181 GTCGGGATGAGGTCATGATACACTTAATCCTCGGCAACAGATTTATTAATATCTTGAT 240
QY 241 GATGAATTAACGAAGTTGATGGTGGAGAAAATATACATCGATTAATATGTCATAAACA 300
Db 241 GAGGAACATGACAGCGCTTTAGGTTCTGATACGGCAGAAATTTATCAAGTCACCTAAGAT 300
QY 301 CCTACTGTTGTTATGATGTTGGTTTACAAGGTGCTGGTAAACACAACTCCAGGTA 360
Db 301 CCAACCATCATCATGATGTTGGTTTACAGGGGCTGGTAAACAACTCTGCTGTAAA 360
QY 361 TTAGCATTTAATGATGCTGTAACAAATAACAAACAAACCTATGTTAGTTCAGCAGATATT 420
Db 361 TTGGCCACAACTCAAGAAAGAGAAATGCTGCTCTTTGATGATTCGGGCGGATATT 420
QY 421 TATCTCCAGCAGGATAAATCAATTAACAAACAGTAGGAAACAAATTAATGATTCCTGTA 480
Db 421 TATCTCCAGCTGCCATTGACCAAGCTTAAGACCTTGGGCAACAGATTTGATGCTGCTGC 480
QY 481 TACAGTGAAGGAGATCAAGTAAAGCCCAACAAATTTGTAACATTAATCCATTAACATGCT 540
Db 481 TTTGCACCTTGAAGAGAGATACCAAGCTGTTGAGATTTAGTCAAGGTTGGAGCAAGCC 540
QY 541 AAAGAAGAACATTTAGACTTTGTAATCATGATACAGAGTGTGATTAACATCATGATGAA 600
Db 541 CAACTAATCATACAGCTATGCTTTGATGATGATGCTGCTGCTTTGATGATTCGGGCGGAT 600
QY 601 GCATTGTAAGCAATTAAGAAAGTAAAGAAATTTGCTAAACCAACAAATTAATGATTTA 660
Db 601 CTCCTCATGAATGAGCTTCTGATGTAAGCAATTTGGCTCAACCAATTAATGATTTA 660
QY 661 GTTGTGCAATTAAGCGGCTCAAGATGCTGCTCAATGTTGCAAGATCTTTTACACATCA 720
Db 661 GTCGTTGATGCTATGATGTTGGTCAGGAAGCAACCAATGTTGCGGCTGAGTTTAACTCTCAG 720
QY 721 CTGTGATGTCACAGGTTTACCTTAACATAATAGATGATACAGTGTGCTGCTGCTGCTGCT 780
Db 721 TTGGAGTACTGGGTCATCTCTACCAAGATTTGATGCGGATGATGCTGCTGCTGCTGCT 780
QY 781 TTATCTATTCGTTGCGTGAACAAACCAATTAATTTGTTGTTGATGATGATAAAGTTA 840
Db 781 CTGCTCTGTTGCTACATTAATGGAACCAATCAAGTTCACTGCTACAGGTGAAAGATTT 840
QY 841 GATGTTTAGAGCTATTCATCTGCAAGCTATGCAATGATGATGATGATGATGATGATGAT 900
Db 841 AGGACATTTGAACCTTCCACCCAGACCGGCTGCTAGCCGCTATCTGTTGTTGTTGTTG 900
QY 901 GTGTTTAACTTTAATTTGAAAGGCAACCAAGATTTGATCAAGAAACCAACCAAGATTTA 960
Db 901 ATGCTCACTTTGATTTGAGAAAGCTTCTCAGGAATACATGAAACCAACCAACCTTTGAA 960
QY 961 GAGAAAAAGATGCGTGAGTCTGCTTTTACTTTAGATGATTTTGTAGAACAACTTGATCAG 1020
Db 961 GCTGAGAAAGATGCGGAAACCACTTTGATTTTAAATGATTTTATGATTTTATGATTTTAT 1020
QY 1021 GTGAAAAATCTAGACCCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 GTGCAAAATATGCGGCGGCTGGAAGACTTGTCTCAAGATGATTTCCAGGTTATGCGCA 1080
QY 1081 AAAGGCTTAGATGATTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 CCAGCATTCAAAACATGAAGTGGTGAAGCCAGATTTGCTCGTAAACGCTGCCATTTG 1140
QY 1141 CAGTCATGACCGCGGCTGAAAGAAACCAATCCAGACACATTTGAATGATCATCAATTA 1200
Db 1141 TCTTGGATGACCTGAGAGACGTGAAACCCAGATTTGTTTAATCCAGCGCTGCGCT 1200

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9156
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1572)
US-09-815-242-9156

Query Match 36.3%; Score 496; DB 10; Length 1572;

Best Local Similarity 61.1%; Pred. No. 2.5e-99;

Matches 802; Conservative 0; Mismatches 510; Indels 0; Gaps 0;

QY 1 ATGGCATTTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAATGGTGGTAAG 60
DB 1 ATGGCATTTGAAGCTTATCAGACGCTGCAAGCGACGATGCAAAAATGGTGGTAAG 60
QY 61 GGTAACTTACTAGCTGATATTAAGATAATGATGCGTGAAGTAAGATACGCTTATTT 120
DB 61 GGTAACTTACTAGCTGATATTAAGATAATGATGCGTGAAGTAAGATACGCTTATTT 120
QY 121 GAGGCTGACGTAACTTTAAAGTGTAAAGAAATTTTAAACACATATCAGACGCGCA 180
DB 121 GAGGCGGACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 181 TTAGCTTCCGATGATGAATCAATTAACACAGGCGCAACAACTTATTAATAATGCTCA 240
DB 181 GTCGGGATGAGGTCATGATACACTTAATCTCGCAACAGATTTATTAATGCTGAT 240
QY 241 GATGAATTAAGAGTGTGATGGTGGAGAAATATATCGATTAAATATGCAAAATAACCA 300
DB 241 GAGGAACGTACAGCGCTTTAGCTGATACGCGAGAAATATATCAAGTCACCTAAGATT 300
QY 301 CCTACTGTTGATGATGCTGTTTACAGAGTGTGTTAAACACAACTCCAGGTAAA 360
DB 301 CCAACCATCATGATGATGCTGTTTACAGAGGCTGTTAAACACAACTCCAGGTAAA 360
QY 361 TTAGCATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 TTGCGCAACACATCAGAAAGAGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 421 TATGCTCCAGCGATTAATCAATTAACAACTAGTGGGAAACAAATGATATTCCTGTA 480
DB 421 TATGCTCCAGCGATTAATCAATTAACAACTAGTGGGAAACAAATGATATTCCTGTA 480
QY 481 TACAGTGAAGAGATCAAGTAAGCCACAAATTTGTAATTAATGCAATTAACATGCT 540
DB 481 TTGCACTTTGAACAGAGATACACGCTGTTGAGATTGATGATGATGATGATGATGAT 540
QY 541 AAAGAAGACATTTAGACTTTGATCAATGATACAGAGCTGATGATGATGATGATGATG 600
DB 541 CAAATATCAATACGACTATGCTTCAATGATGATGATGATGATGATGATGATGATG 600
QY 601 GCATTGATGACGAATTTAAAGAGTAAAGAAATTTGCTAAACCAACGAATTTATGTTA 660
DB 601 CTCCTCATGATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 660

QY 661 GTTGTGATTAATCAAGTCAAGCTGCTCAATGCTGCAATCTTTTACAGATCAA 720
DB 661 GTTGTGATTAATCAAGTCAAGCTGCTCAATGCTGCAATCTTTTACAGATCAA 720
QY 721 CTTGATGTCACAGGTGTTACCTTAATTAATAGATGCTGATACACCTGGTGGTCACT 780
DB 721 TTGGAAGTGAAGTGGGTCATCTTACCAAGATGATGCGGATCTGCTGGTGGTCTG 780
QY 781 TTAATCTATGCTGGTGGTGAACAAACCAATTAATTTTGGTATGAGTGAAGTTA 840
DB 781 CTTGCTGTTGCTGATCAATTTGGAACCAATTAATTTTGGTATGAGTGAAGTTA 840
QY 841 GATGTTTATGAGCTATTCATCTGATGATGATGATGATGATGATGATGATGATG 900
DB 841 AGGACATTTGAACCTTCCACCCAGACCGCATGCTAGCCGCTATCTTGGTATGGG 900
QY 901 GTGTTAATTTAATGAAAAAGCGCAACAGATGCTGATCAAGAAAAAGCAAAAGATT 960
DB 901 ATGCTCACTTGTGATGAGAAAGCTTCTCAGGAATACGATGACAAAAAGCCCTTGA 960
QY 961 GAGAAAAAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 961 GCTGAGAAAGATGCGCGAAAAACACCTTTGATTTAATGATTTATGATTCATCAT 1020
QY 1021 GTGAAAAATCTAGACACCTGATGATGATGATGATGATGATGATGATGATGATG 1080
DB 1021 GTGAAAAATCTAGACACCTGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1081 AAAGGCTTAGATAGCTTAAATGATGATGATGATGATGATGATGATGATGATGATG 1140
DB 1081 CAGGCGCTTCAAAACATGAGGTGGATGAGCCAGATGCTCGTAAAGCTGCCATGT 1140
QY 1141 CAGTCAATGAGCGCGGTGAAGAAACAAATCCAGACATGATGATGATGATGATGAT 1200
DB 1141 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 CTTATGCTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 1201 CTTATGCTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1261 AACCATATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312
DB 1261 AACCATATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1312

RESULT 13

US-09-974-300-1847
; Sequence 1847, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1847
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1847

Query Match 35.6%; Score 486.4; DB 10; Length 1136;

Best Local Similarity 64.3%; Pred. No. 2.8e-97;

Matches 730; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 1 ATGGCATTGTAAGGCTTATCAGAACGCTCGAAGCAGCAGTCAAAAATTCGCGTGAAG 60
Db 1 ATGGCATTGTAAGGCTTATCAGAACGCTCGAAGCAGCAGTCAAAAATTCGCGTGAAG 60
QY 61 GGTAAATTTACTGAGTGAATATAAGATAATATGTCGCTGAAGTAAGATTAGCGTTATTT 120
Db 61 GGAAGGTAACGAGCAGCAGTCAAGAGATGATGCGGAGTCCGTTAGCGCTCCTT 120
QY 121 GAGCTGACGTAACTTTAAAGTGTAAAGATTTATTAACACAGTATCAGAACCGCA 180
Db 121 GAAGCGAGCTCAATTTTAAAGTCGTGAAGATTTGTGAAAAAGGTAAGCGACAGCT 180
QY 181 TTAGGTTCCGATGTAATCAATCATTAACACAGGCGCAAGTATTAAAAATAGTTCAA 240
Db 181 GTGGACAGAGCTTATGAAAAGCCTGACGCCGACAGCAGTCAATTAAGTGTAAAC 240
QY 241 GATGAATTAACGAATTTGATGGGTGGAGAAAATPACATGATTAATATGTCARATAACCA 300
Db 241 GAAGAGTCAAGGAGCTGATGGCGGCGAGAAACAGATTCGCGTGCAAAAGCGCTCCG 300
QY 301 CCTACTGTTTATGATGTTGGTTTACAGGCTGCTGGTAAACCAACACTGCAGGTAA 360
Db 301 CGACCGCTCATCATGCTGGCTGTCGAAGGGCGCGTAAACAGACGTGCGCCGGAAG 360
QY 361 TTAGCATTATTGATCGTAAAAATACACAAAAAACCTTATGTTAGTTCAGCAGATAT 420
Db 361 CTGGCAACCTGCTCGCAAAAACATACCGCAATCCGCTGCTGTCGTCGAGCAGATC 420
QY 421 TATGCTCCAGCAGGATAAATCAATTAACAAACAGTAGGGAACAAATGATATTCCTGTA 480
Db 421 TACGCCCGCGCGCATCAACAGCTGGAACGCTCGCAAGCAGCTCGATATGCGGTA 480
QY 481 TACAGTGAAGGAGATCAAGTAAACCCACAAATTTGAATTAATGCAATTAACACATGCT 540
Db 481 TTTTCACTGGGTGACAGGCTCAGTCCGGTGGAGATTGCCAAGCGGCATCGAAAGGCG 540
QY 541 AAAGAAGACATTATGACTTTGTAATCAATGATACAGCAGGTCGATTACACATCGATGA 600
Db 541 AAAGAGATCATCAGACCTAGCTTCTCATGATACGGCGGACGCTTCATATGATGA 600
QY 601 GCATGATGACGAATTAAGAGAGTAAAGAAATTTGCTAAACCAAGAAATTTATGTTA 660
Db 601 GAGCTGATGATGAGCTTGAGCAGGTAAGAAACCGCTCAGCGGGAAGAGATTTCCTT 660
QY 661 GTTCTGATTCATGACGGGTCAGATGCTGTCAATGTTGCAAGATCTTTTGACGATCAA 720
Db 661 GTCTGACTCCATGACGGGTCAAGACGCGGTCAATGTTSCCAAAGCTTTTAAGCAAC 720
QY 721 CTTGATGTCACAGGCTTACCTTAACATAATAGATGGTGATACACGTTGGTGGAGCT 780
Db 721 CTCGGTGTACAGGTGCTCATCTTGACGAGCTTGACGGCGATCTAGAGCGGAGCGGC 780
QY 781 TTAATCATTCGTTGCTGACAAACCAATTAATTTGTTGGTATGAGTGAAAGTTA 840
Db 781 TTGTCAATTCGCGCGCTGACAAACACGCGCATCAAAATGCGCGGAATGGCGAAAGCTT 840
QY 841 GATGGTTTACAGCTATTCCTTCAGCTGATGCATCAGTATTTTAGTGTAGTGGTAT 900
Db 841 GATCGCTGACGCGCTTTCATCTGACGCGATGGCTTCGCGGATCTCGGAATGGCGAT 900
QY 901 GTGTTAAGTTTAAATGAAAACGCAACAGATGTGTGATCAAGAAAACCAAGATTTA 960
Db 901 GTGCTACCTTATCGAAGAACCCGACGCCAAGCTAGACGAGAAAGCGAGGAATC 960
QY 961 GAGAAAAGATGCTGAGTCAATCGTTTACTTTAGATGATTTTTTGAACAACCTTGATC 1020
Db 961 GAGCAAAAANTGAAAATTAAGCTTTTACGCTGACGACTTTTGAACAGCTCGGACAG 1020
QY 1021 GTGAAAATATGACACCTCGATGATTAATTAAGAAATGATTCAGGATGATGAATAATG 1080
Db 1021 GTCCGCAATATGGGCGCGCTTGAAGACCTGATTCAAATGATGTCGCGGCGCAAAATG 1080
QY 1081 AAGGGCTAGNTAGCTTAATATGATGTAAGTCAAGAAACAAATTTGATCATAATTAAGCGAT 1136

Db 1081 AAAGGCTTAAAAAAGCTGAAGGTTAGTAAACACAGCTCAGCCATATCGAAGCAT 1136
RESULT 14
US-09-815-242-3070/c
; Sequence 3070, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3070
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-3070
Query Match 33.9%; Score 463.2; DB 10; Length 468;
Best Local Similarity 99.4%; Pred. No. 2.6e-92;
Matches 465; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 103 GTAAGATTAGCGTTATTTGAGGCTGACGTAAACTTTTAAAGTGGTAAAAAGAAATTTATTTAA 162
Db 468 GTAAGATTAGCGTTACTTTGAGGCTGACGTAAACTTTTAAAGTGGTAAAAAGAAATTTATTTAA 409
QY 163 ACAGTATCAGAACGCGCATTAGGTTCCGATGTAATGCAATCATTAACACCGGCGACAA 222
Db 408 ACAGTATCAGAACGCGCATTAGGTTCCGATGTAATGCAATCATTAACACCGGCGACAA 349
QY 223 GTTATTAATAATAGTTCACAGATGAATTAACAGATGATGGTGGAGAAAATACATCGATT 282
Db 348 GTTATTAATAATAGTTCACAGATGAATTAACAGATGATGGTGGAGAAAATACATCGATT 289
QY 283 AATATGCAAAATAAACCCACTACTGTTTATGATGCTTGGTTTACAAAGGTCGTGTTAA 342
Db 288 AATATGCAAAATAAACCCACTACTGTTTATGATGCTTGGTTTACAAAGGTCGTGTTAA 229
QY 343 ACAACAACCTCGAGGTAATTAGCATTTAGCATTTAGTGGTAAATAATACAAACAAAACCTATG 402
Db 228 ACAACAACCTCGAGGTAATTAGCATTTAGCATTTAGTGGTAAATAATACAAACAAAACCTATG 169
QY 403 TTAGTTCACAGCATATTTATCGTCCACGCGCATTAATCAATTTACAAACAGTAGGAGAA 462
Db 168 TTAGTTCACAGCATATTTATCGTCCACGCGCATTAATCAATTTACAAACAGTAGGAGAA 109
QY 463 CAAATTGATATTCTCTGATACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAAC 522

```
Db 108 CAAATGGATATTCCTGTATACGTGAAGGAGATCAAGTAAAGCCACACAAATTGTAAC 49
QY 523 AATGCATTAACACATGCTAAAGAACACATTTAGACTTTGTAATCATT 570
Db 48 AATGCATTAACACATGCTAAAGAACACATTTAGACTTTGTAATCATT 1
RESULT 15
US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ IDS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868
Query Match 32.4%; Score 443.4; DB 10; Length 1389;
Best Local Similarity 59.2%; Pred. No. 7.2e-88;
Matches 775; Conservative 0; Mismatches 531; Indels 3; Gaps 1;
QY 7 TTTGAGGCTTATCAGACGCGCTCCAGCGACGATGCCAAAATTCGCTGTAAGGTAA 66
Db 4 TTTGAGATTTTATCGGATCGCGCTTTCCAAAACCTTACGTAATATCACAGAAAAGCGCT 63
QY 67 CTTACTGAGGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTATTGAGGCT 126
Db 64 TTACCGAGAGATAATTAAGAAACCTTACGCGGAGTGCCTATGCGATTAAGGCT 123
QY 127 GAGCTAAACTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGCGCATTAGT 186
Db 124 GATGTTGCCCTGCTGCTGCGTGAATTTATCGCAAAAAGTAAAGAAAGCGCGCTGGG 183
QY 187 TCCGATGTAATCATCATCATTAACACAGGCGCACACAGTTATTAATAGTTCAAGTGA 246
Db 184 GAAGAAGTCAATTAAGAAGTTTAACCCGAGGCAAGATTCTTAAAAATCGTTACAGCGT 243
QY 247 TTAACGAAGTTGATGGTGGAGAAAATACATCATGATTAATGTCAAAATAAACCACTACT 306
```

```
Db 244 CTTGAAAAGCCATGGCGAAGCGAATGAGAGTTTAAATCTCGCAACCCACACACGCA 303
QY 307 GTTGTATGATGGTGTGTTTACAAAGCTGCTGTAATAACACAACTGCAGGTAAATTAGCA 366
Db 304 GTTATCTTAATGCGCGGTTTACAAAGGCGGTAAACACCACGAGTGTGGTAAATGGCA 363
QY 367 TTATGTAGTGGGTAAATAATACAAACAAAACCTATGTTAGTTGTGACAGATATTTATCGT 426
Db 364 AAATTTCTTGGCTGAACGCCATAAAAGAAAGTGTAGTGGTGTCTGCTGACGTATATGCG 423
QY 427 CCAGCAGCGATTAATCAATTACAAACAGTAGGAAACAAATTTGATATTCTCTATATACGT 486
Db 424 CTTGCTGGGATTAAGCAACTTGAACCTTGGCTCAATCCGTTGGCGTGGATTTTTCACA 483
QY 487 GAAGGAGATCAAGTAAAGCCACACAAATTTGAACATAATCAATTAATAACATGCTAAAGAA 546
Db 484 TCGGATGTTAAACAAACCCCTTGATATTGCTAAATCGCGGCTGCTGATGCAAAACTG 543
QY 547 GAACATTTAGACTTTGTAATCATTTAGTACAGAGTGTGATTAACATCATCATGATGAAGCAT 606
Db 544 AAATTTCTAGATGTTGATTTGATTTGATGATGATGATGATGATGATGATGATGATGATG 603
QY 607 ATGACGATTAAGAGAGTAAAGAAATTTGCTAAACCAACGAAATTAATGTTAGTTGTC 666
Db 604 ATGGAGAAATCAAGCAAGTCCATGCTGCATTAATCAATCAATCAATCAATCAATCAAT 663
QY 667 GATTCATGACGGGTCAAGATGCTGTCAATGTTGCAGAAATTTTGGACGATCAACTTGTAT 726
Db 664 GATGCGATGACTGTTCAAGATGCGCAATATACAGCAAAAGCGCTTATGAAGCATGCTT 723
QY 727 GTCACAGGTGTTTACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 786
Db 724 CTTACAGGGGTTATTTTACGAAAAGTGGACGCTGATGCGCGCGGTGTCGCGCTTATCG 783
QY 787 ATTCGCTGGTGCACAAAAACCAATTTAAATTTGTTGATGATGATGATGATGATGATGAT 846
Db 784 ATTCTGCAATCACAGGTAAACCAATCAATTTCTTGGGTGGCGGAGAAACAGAGCGG 843
QY 847 TTAGACTATTCCATCTCTGAACGATGGCATCACTATTTTATGATGATGATGATGATGATG 906
Db 844 CTTGAGCCATTCATTCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
QY 907 AGTTTAATTAAGAAAGCGCACACAGATGGATCAAGAAAGCAAAAGATTTAG--AG 963
Db 904 TCCCTTATCGAAGATCTTGAACGCTTCTGTGATGATGATGATGATGATGATGATGATG 963
QY 964 AAAAGATGCGGTGAGTCACTGTTTACTTTAGTATGATTTTATGAAACACTTGATCAGGTG 1023
Db 964 AAATTCAGAAAGCGGATGATTTTACTTTAGATGATTTTCCGCGACACAGCTGATCGAA 1023
QY 1024 AAAATCTAGGACCACTGGATGATATTAAGAAATGATTCACAGTATGATTAATAAGTAAA 1083
Db 1024 AAAAATGGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1083
QY 1084 GGCTAGATGAAGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
Db 1084 GAGCAGTTTAAATAATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
QY 1144 TCAATGACCCCGCTGAAGAAACAAATCCAGACACATTTGAATGATCATCGTAAAGACGT 1203
Db 1144 TCCATGACCTTAAAGAACGCTGCCACCCAGATATTAACAAGGATCTCGCGCTCGCT 1203
QY 1204 ATTGCTAAAGGCTGCGGTGCTGATTAATTAAGAAAGTCAATCGTTTGAAGAAACATTTAC 1263
Db 1204 ATTGCTAATAGGCTGCGCACTCAAGTCAAGATGATGATGATGATGATGATGATGATGAT 1263
QY 1264 GATATGAAGAAATGATGAACAAATTCACCTGCTGCGGTAAAGGTAAA 1312
Db 1264 GAAATGCAAGTATGATGAAGAAATTCGCTGAAAGCGGCGATGGCTAAA 1312
```

• • • • •

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2003, 21:34:56 ; Search time 1861 Seconds
(without alignments)
11905.119 Million cell updates/sec

Title: US-09-943-108A-1

Perfect score: 1368

Sequence: 1 atggcattgaagcgtatc.....gtatgaattaccgttttaa 1368

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em.estba:*

2: em.esthum:*

3: em.estin:*

4: em.estnu:*

5: em.estov:*

6: em.estpl:*

7: em.estro:*

8: em.htc:*

9: gb.est1:*

10: gb.est2:*

11: gb.htc:*

12: gb.est3:*

13: gb.est4:*

14: gb.est5:*

15: em.estfun:*

16: em.estom:*

17: gb.gss:*

18: em.gss.hum:*

19: em.gss.inv:*

20: em.gss.pln:*

21: em.gss.vrt:*

22: em.gss.fun:*

23: em.gss.mam:*

24: em.gss.mus:*

25: em.gss.other:*

26: em.gss.pro:*

27: em.gss.rod:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	422	30.8	2169	17	BH770912 LLMGtag64
c 2	248.6	18.2	773	13	BF934118 EST554007
c 3	243.8	17.8	711	12	BF940797 7G99A06.x
c 4	232.2	17.0	875	17	AF075878 AF075878
c 5	223	16.3	656	14	B0805434 B0805434
c 6	221.2	16.2	593	10	AM648731 EST327101

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	217.2	15.9	688	17	BH394904
8	210.4	15.4	677	13	EM413418
9	190.6	13.9	605	10	BE432304
10	187.6	13.7	514	13	BM323434
11	186.4	13.6	581	9	A1977909
12	183.4	13.4	814	17	A2529841
13	179	13.1	1200	11	AY109321
14	178.8	13.1	677	14	BQ96389
15	177.4	13.0	533	10	AW930742
c 16	176.2	12.9	889	17	AZ682189
c 17	163.8	12.0	891	17	BH147501
18	163.2	11.9	2387	11	AK011928
19	158.2	11.6	465	9	AI087737
20	154.2	11.3	554	17	BH369841
21	149.8	11.0	490	13	BM325805
22	148.2	10.8	1032	11	AY106667
23	147	10.7	483	10	BE596266
24	138.6	10.1	476	10	AW033339
25	135.6	9.9	382	10	AW649149
26	133	9.9	1020	17	CNS07B0X
27	131.6	9.6	542	13	BF12818
28	127.8	9.3	850	10	AW147928
29	122.6	9.0	899	14	BQ736950
30	120.2	8.8	773	13	BF454259
31	119.6	8.7	703	12	BF936572
32	119	8.7	573	12	EG317280
33	118.6	8.7	655	14	BQ998609
34	118.6	8.7	656	12	BF641333
35	118.6	8.7	1046	13	BF948099
c 36	117.6	8.6	396	9	AJ500175
37	117.6	8.6	791	17	AQ327477
38	117.4	8.6	553	12	BG315860
39	116.6	8.5	307	13	BF307701
c 40	115.8	8.5	481	17	BH376496
41	115	8.4	874	12	BG679799
42	114	8.3	610	14	BM944643
43	112.2	8.2	666	9	AI726938
44	112	8.2	547	10	BE461473
45	112	8.2	639	14	BQ608517

ALIGNMENTS

RESULT 1
BH770912/c
LOCUS
DEFINITION
Lactococcus lactis subsp. cremoris genomic, DNA sequence.
ACCESSION
BH770912
VERSION
BH770912.1
KEYWORDS
GSS.
SOURCE
Lactococcus lactis subsp. cremoris
ORGANISM
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
REFERENCE
1 (bases 1 to 2169)
Bolotin,A., Ehrlich,S.D. and Sorokin,A.
TITLE
Studies of genomes of dairy bacteria Lactococcus lactis
JOURNAL
Sci. Aliments, (2002) In press
COMMENT
Contact: Sorokin A
Genetique Microbienne
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL1403 is ffh (91%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 2141.
Location/Qualifiers
l. .2169

BH770912 2169 bp DNA linear GSS 01-MAY-2002
LLMGtag644 MG1363 Random Sequence Tag Library Lactococcus lactis
subsp. cremoris genomic, DNA sequence.

ACCESSION BH770912.1 GI:20373869

VERSION BH770912.1

KEYWORDS GSS.

SOURCE Lactococcus lactis subsp. cremoris.

ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Lactococcus.

REFERENCE 1 (bases 1 to 2169)

Bolotin,A., Ehrlich,S.D. and Sorokin,A.

TITLE Studies of genomes of dairy bacteria Lactococcus lactis

JOURNAL Sci. Aliments, (2002) In press

COMMENT Contact: Sorokin A

Genetique Microbienne

INRA

CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Tel: 33 1 34 65 25 16

Fax: 33 1 34 65 25 21

Email: sorokine@jouy.inra.fr

best homologue in strain IL1403 is ffh (91%)

Class: shotgun

High quality sequence start: 30

High quality sequence stop: 2141.

Location/Qualifiers

l. .2169

```

/organism="Lactococcus lactis subsp. cremoris"
/strain="MGI363"
/db_xref="taxon:1359"
/clone_lib="MGI363 Random Sequence Tag Library"
/notes="Vector: pSGM2; Site_1: SmaI; library of
chromosomal fragments of L.lactis strain MGI363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT      642 a   429 c   380 g   718 t
ORIGIN
Query Match      30.8%; Score 422; DB 17; Length 2169;
Best Local Similarity 63.7%; Pred. No. 1.4e-80;
Matches 689; Conservative 0; Mismatches 390; Indels 3; Gaps 3;
QY 225 TATTAAATAGTTCAGATGAATTAACGAGTTGATGGTGGAGAAATACATFCGATTA 284
DB 2169 TATCAAAATTTGTATGAAGAATTAACAATATCTTAGCGGTGTGTGAAGCCGAATTA 2110
QY 285 TATGTCAAATAAACCACTACTGTGTATGAT-GGTTGGTTTACAGGTGCTGTGTA 343
DB 2109 CAATCTCTAAATCCCAACAATATCATGATGGGTGGTTTGCAGGGGCGAGGTAA 2050
QY 344 CAACAACCTCAGGTAATATAGCATATATGATCGTAAATAATACACAAAACCTATG- 402
DB 2049 CAACCTTTCGTGTAACTTCGCAAAAACCTTAAAGAAAGAACAAATGCTCGCCCACTGA 1990
QY 403 TTAGTTGCAGCAGATATTTATCCGCCAGCAGCA-TAATCAATTAACAACAGTAGGAA 461
DB 1989 TGATTTGGCAGCGAGCTTTATCGCCAGCAGCATTTGGACCAATTAATAAACCTTGGGAGA 1930
QY 462 ACRAATTTGATATTCCTGTATACAGTAGGAGGATCAAGTAAGCCACACAAATTTGTAAC 521
DB 1929 ACRACTTGACATTCCTGTTTATGATGAGGAGAGCTGTGAAACCCGTTAATATTGTCG 1870
QY 522 TAAATGATTAACAACATGCTTAAGAGAACATTTAGACTTTGTAATCATTAATGAACAGCAG 581
DB 1869 TAATGTCCTTTTAAAGCTTCRAAGAACGTTAAAGATTTATGTTGATTGATAGGCGAG 1810
QY 582 TCATTAACATCGATCGATGAGCATTTGATGAACGAATTAAGAAGTAAAGAAATTCCTAA 641
DB 1809 TCGCTTGAATGATGATACATTTGATGATGAGCTTCAAGAGATTAACAACCTCGCCA 1750
QY 642 ACCAAACGAATATGTTAGTTGTCGATTCATATGACGGGTCAAGATGCTCAATGTTGC 701
DB 1749 ACCAACAGAAATTTTACCTGTTGTGATGCGATGACTGCTCAAGTGCACCAAGTGC 1690
QY 702 AGAATCTTTTGACGATCACTTGATGTCACAGGTGTACTTAACTAAATTAGATGGTGA 761
DB 1689 TAAACTTTTGATGAATAATAGATATATACCGGTGTATCATCACCACATCGATGGGA 1630
QY 762 TACACGFGTGGTGCAGCTTATCTATTCGTTGGTGAACAAAACCAATTAATTTGT 821
DB 1629 CACACGFGTGGTGGCGTTTATCAATTCGTGAATTAACCGGAACCACTTAAATTTAC 1570
QY 822 TGGTATGAGTGAAGATAGATGTTTAGAGCTATTCATCTCTGAAGTATGCATCAG 881
DB 1569 AGGACTGGTGAATAATTAACCGACCTTGAAGTTTCTACCTCATGATGATGAGCTCAG 1510
QY 882 TATTTAGGTATGGTGTGTTTAACTTTAATTAAGAAAGCGCAACAGATGTGATCA 941
DB 1509 AATCTGGTATGGGTGATGCTCACTCTGATTGAGAGGCGCAAGCTAATATGATCA 1450
QY 942 AGAAAAAGCAAGATTTAGAGAAAAGATCGCTGAGTCATCCTTTACTTATGATGATT 1001
DB 1449 GGAACATCTGCAAAACTCGCTGAATAATGGCTGATTAACCGTTTGGATTAAGGATTT 1390
QY 1002 TTTAGAACAACTGATCAGGTGAATAATCTAGGACCATGATGATATATGAAATGAT 1061
DB 1389 GTTTGACCACTAGATCAAGTTTACAAATATGGGACCATGGAAGATATATGAAATGAT 1330
QY 1062 TCCAGGTATGAATAAAGAGGCTAGATAAGCTTAATATGATGAGTGAAGAAATTTGA 1121
DB 1329 TCCAGGAATGTCAAAATSCCTGGCTTGGATAAAGTTAAAGTTGACCCCTAAAGATGTTGC 1270

```

```

QY 1122 TCATTTAAAGCGGATTAATCAATCAATGACCGCGCTGAAGAAACAATCCAGACACATT 1181
DB 1269 TAGAAAACGAGCAATGATTTGTCATGTAATGCCCGCAGCTCAATTTAGAGCAGAACT 1210
QY 1182 GATGTATCACTGTAAGAAAGCGGATTTGCTTAAGAGGTGCTGCTTCATTACAGAAATCAA 1241
DB 1209 ATCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
QY 1242 TCGTTTGTGATGAACAATTTAAGCATATGAGAAAATGATGAACAATTCAGTGGTGGCG 1301
DB 1149 TAAATTCATTAAAGCAATTTAATCAATCCAAAGAAATGATCAAGGAATCAATGAATGTGA 1090
QY 1302 TA 1303
DB 1089 TA 1088

RESULT 2
BI934118
LOCUS      BI934118
DEFINITION EST554007 tomato flower, anthesis Lycopersicon esculentum cDNA
ACCESSION  BI934118
VERSION     BI934118.1 GI:16248590
KEYWORDS   EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE   1 (bases 1 to 773)
AUTHORS     van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
            Uterback,I., Van Aken,S., Renning,C.M., Nierman,W., Fraser,C.M.,
            Martlb,G.B., Giovannoni,J.J. and Tanksley,S.D.
            Generation of ESTs from tomato flower tissue, anthesis (2001)
            Unpublished (2001)
            Contact: CUGI
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source            1..773
     organism="Lycopersicon esculentum"
     cultivar="TA496"
     db_xref="taxon:4081"
     clone_lib="tomato flower, anthesis"
     clone_lib="tomato flower, anthesis"
     tissue_type="flower"
     dev_stage="anthesis"
     note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; supplier: Cornell University; sequencing: the
            Institute for Genomic Research; Flower buds and flowers
            were taken from greenhouse plants (4-8 wks old, TA496).
            They were immediately frozen in liquid nitrogen and then
            size-separated while remaining frozen."
BASE COUNT      227 a   129 c   210 g   207 t
ORIGIN
Query Match      18.2%; Score 248.6; DB 13; Length 773;
Best Local Similarity 58.5%; Pred. No. 2.5e-43;
Matches 453; Conservative 0; Mismatches 319; Indels 3; Gaps 1;
QY 116 TATTTCAGCTGACGTAAATTTTAAAGTGGTAAAGATTTATTAACAGTATCAGAAC 175
DB 2 TTTTGAAGCTGATGTTAGTCTCCCGAGTTGTCAGAGGTTTGTTCAGTCTGTAGTAGAG 61
QY 176 GCGATTAGTTCGCGATTAATGCAATCATTAAACACGAGGCAACAGTTATTAAATAG 235

```

```

Db 62 AAGCGTTGGCACTGTTGATTCAGAGTGTAAACCCAGATCAGCAACTAGTTRAGACT 121
QY 236 TTCAGATGAATTAACGAAGTTGATGCGGTGAGAAATACATCGATTAATATGTCRAATA 295
Db 122 TACGTGAGAGCTGTGAACACTGATGGGTGAGAGGTTTCTGAAGTGGTTTTTGTCTAAAT 181
QY 296 AACCACCTACTGTGTTGATGATGTTGTTTACAAAGTGTGTTTAAACCAACACTCCAG 355
Db 182 CTAAGCCACAGTAATACATATTGCGCGCTACAAAGCTGTGGAGACACTGTTAGT 241
QY 356 GTAATATAGCATATGATGCGGTAAATAATACAAACAAACCTATGTTAGTCCACAG 415
Db 242 CAAAGTTAGCTTTAT---ATCTAAAGAACAGATGTAAGATTGATGTTGCTGGG 298
QY 416 ATATTATCGCCACAGGATAATCAATCAACACAGTAGGAAACAAATTCATATTC 475
Db 299 AGTGTACAGACCTGCTGCTATGACCACTGTTATTTGGGTAAACAGGTTGATGTAC 358
QY 476 CTGTATACAGTAAGAGATCAAGTAAGCCACACAAATTTTAACTAATCATTAAC 535
Db 359 CTGTTATGACAGGAGACAGATGTAACAACTGCAGAAATAGCCGACAAAGATTACAAG 418
QY 536 ATGCTAAGAGACATTTAGACTTTGTAATCATTTGATACAGAGTGCATTACATCG 595
Db 419 AGCCCAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 478
QY 596 ATGAGCAATTCATGACAGGATTAAGAGAGTAAGAAATTTGCTAAACCAACGAAATTA 655
Db 479 ATAAACTATGATGATGATTAAGAGAGTGAACGGGTACTGAACCCACAGAGTTT 538
QY 656 TCTTAGTTGCGATCAATGACGGGTCAAGATGCTGCAATGTCAGAACTTTTTCAGG 715
Db 539 TACTTTGTTGGATGCAATGATGCTGCAAGAGCTGAGCTTTGCTCAACATTCATC 598
QY 716 ATCAACTTGTATGTCAGAGTGTGTTTAACTTAATTTAGATGTTGATACAGCTGGTG 775
Db 599 TCGAATGAGGATGATGCTGCTGCTATGTCAGAGCTGATGAGATTCAGGGGTGAG 658
QY 776 CAGCTTTATCATTCATGCTGGGTGACAAACAAACCAATTAATTTGTTGATGATGAGAA 835
Db 659 CAGCTTTTATGTCAGAGGATGATGAGGAAAGCAATCAAGCTCGTAGGAGGGGTGAAC 718
QY 836 AGTTAGATGGTTTATGATGATTCCTGCTGAGCTGTCAGCTGTCAGCTATTTTAG 890
Db 719 GTATGAGGACCTTGAACTTCTATCTGCTGACCGCATGCTGCTGATGATTTTAGG 773

RESULT 3
BF940797/c 711 bp mRNA linear EST 22-JAN-2001
LOCUS 7439a06.x1 lupski_dorsal_root_ganglion Homo sapiens cDNA clone
DEFINITION IMAGE:13280811.3 similar to SW:SR54_BAC50 P37105 SIGNAL RECOGNITION
PARTICLE PROTEIN ;, mRNA sequence.
ACCESSION BF940797
VERSION BF940797.1 GI:12358117
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 711)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgap@rs-reimail.nih.gov
Tissue Procurement: Dr. James Lupski
cDNA Library Preparation: Lupski Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 433.
Location/Qualifiers
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="lupski_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
Noti; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACATGTTAGATCGCGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 193 a 145 c 88 g 283 t 2 others
ORIGIN

```

```

Query Match 17.8%; Score 243.8; DB 12; Length 711;
Best Local Similarity 64.2%; Pred. No. 2.7e-42;
Matches 415; Conservative 0; Mismatches 223; Indels 8; Gaps 3;
QY 604 TTGATGAACGAATTTAAAGAGTAAGAAATGCTAAACCAACGAAATTAATGTTAGTT 663
Db 708 TTAATGCCAGAGTAAAGATGCTAAAGAGTTTCAAAATCCTAGTGAATATTATTAGTA 649
QY 664 GTGATTCATGACGGGTCAAGATGCTCAATGTGCGAGATCTTTTACGATCAACTT 723
Db 648 GTTGATGCTATGACAGTCCAGATGCTGTTAAGTAGCAGACATTTAATACTCAATTA 589
QY 724 GATGTCACAG-GTGTTACCTTAATAATTTAGATGCTGATACAGCTGGTGGTCACTTT 782
Db 588 GATCTTTTCAGNGAATAATATTAAACAAAGTTTAGATGGTGATACAAAGAGTGGTCTCT 529
QY 793 ATCTATTCGTTGGTGACACAAACCAATTAATTTGTTGGTATGATGAAAGTAGA 842
Db 528 TTCAATTAAGAGACATTAAGTGGCAAGCCAAATTAATTTGTTGGTGGTGAACAGATG 469
QY 843 TGGTTTAGAGCTATTCCATCCTGAACTGATGCGCATCAGCTATTTTAGTGGTGATGT 902
Db 468 TGATATTTGAAGTATCCATCCAGATAGATGCTTCAAGATATTAGAGATGGAGATCT 409
QY 903 GTTAAGTTTAATGAAAAGGCGCAAGATGTTGATCAAGAAACCAACAAAGATTAGA 962
Db 408 CCTTTCTTTAATAGAAAAGCTCAGCAAGCTATTGACCAAGATGATTCAGAGTGTAA 349
QY 963 GAAAAGATGCGGTGATGCTATCTTTTACTTTAGATGATTTTGTAGAACAACTTATCAT 1022
Db 348 TGAAAAGATGTTAAATCAAGAAATTTTACTTTGATGACTACTTATCAGCTATGATCA 289
QY 1023 GAAAATCTAGGACCACTGGATGATATTATGAAATGATTCAGAGTATGAAATGAA 1082
Db 288 GAAAAGCTGGACCTATAAATAATGATGATGATGATTCAGAGTGTAAACAAAGAAC 229
QY 1083 AGGCTGAGA-TAAGCTTAATATGATGATGATGATGATGATGATGATGATGATGATG 1141
Db 228 TTGAAGGCATTTGATTTTCTCAAGCAGAAAACAAATGCTACAGTTAAAGCAATCATAC 169
QY 1142 AGTCAATGACCGCGGTGAAGAAACCAATCC-----AGACACATTTGAATGATCACGTA 1195
Db 168 AATCGATGACAGCTAAAGAGAGAAACAAACCTTCTTTAGTATAGGAATGTTTCAGAA 109
QY 1196 AAAAGCGATTTGCTAAAGGGTCTGGTCTTCATTAAGAAAGTCAA 1241

```

```

Db 108 AGAGAGATAGCTAAAGTTCTGGTACACAGTACAGAGNTAA 63

RESULT 4
LOCUS AF075878/c 875 bp DNA linear GSS 29-AUG-2000
DEFINITION AF075878 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 245-T3, DNA sequence.
ACCESSION AF075878
VERSION AF075878.1 GI:3320748
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
1 (bases 1 to 875)
Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
Sample sequencing of a Salmonella typhimurium LT2 lambda library:
comparison to the Escherichia coli K12 genome
FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
99243757
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdsu.edu
Class: shotgun.

FEATURES             source
    source
        Location/Qualifiers
            1..875
                /organism="Salmonella typhimurium"
                /strain="LT2"
                /db_xref="taxon:602"
                /clone="245-T3"
                /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                /note="Vector: Lambda DASH II; sequenced using Li-Cor
                sequencer"
BASE COUNT 178 a 266 c 217 g 214 t

Query Match 17.08; Score 232.2; DB 17; Length 875;
Best Local Similarity 56.08; Pred. No. 8.4e-40;
Matches 460; Conservative 0; Mismatches 358; Indels 3; Gaps 1;

QY 296 AACCACTACTGTTGTTATGATGGTTTACAAGGCTGCTGTAACCAACCAACTGCAG 355
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 AGCGCCAGCGGTATGATGCGGGTCTGCGAGGGGGTAAACCAACCGAGGTGG 772
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 GTAAATAGCATATGATGCGTAAATAATACACAAAACCTATGTTAGTTCAGCAG 415
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 GTAAATGCGGTAAATCTTGGCGGAGAGACACAAAGAAAGTGTGCTGCTGCGG 712
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 ATATTTATCGCCAGCAGCAATAATCAATTACAACAGTAGGGAACAAATTTGATATTC 475
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AGTCTATCGCCGGCGGATCAACAGCTCGAAGCGTGGCTGACGAGGTGGCGTGG 652
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 CTGTATACAGTAGAAGAGATCAATAAGCCACACAAATGTTAACTAATGCAATTAAC 535
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 ATTTCTCCGGTCTGATCGGCCAGAAACCGGTTGATGCTCAACCGCGCTGAAAG 592
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 ATGCTAAAGAAGACATTTAGACTTTGATATGATGATACAGCAGGTCGATTACATCG 595
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 AAGGAGCTCAATTTACAGCGTGTGCTGTTGGATACCGCGGTCGCTGCATGTTG 532
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 596 ATGAAGCATTGTAGCAAGTAATAAAGAGTAAAGAAATGCTTAAACCAACCAATTA 655
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ACGAAGCGATGAGGACGAATCAACACAGCTCCACGCTTCATCAACGCCAGTAGAAC 472
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 TGTTAGTTGCAATTCATGACGGGTCAAGATGCTGTCATATGTCGACATCTTTTACG 715
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 TGTTCGTCGATGATGACCGGTCAGGATCGCGGGAATACCGCAAAAGCCTTTAAG 412
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 716 ATCACTTGATGTCACAGGTTTACCTTAATTAATAGATGGGTACACAGTGGTGG 775
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 411 AACGCTGCCGTTAACCGCGTGTGCTGTCACCAAGTGTGATGTGCGCGTGGCGGTG 352
QY 776 CAGCTTTATCTATTTCGTTTCGGTGACACAAAAACCAATTAATTTGTTGGTATGAGTGA 835
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 CGCGGCTCTATTTCGTCATATCACCGGCAAGCGGATTAATTCCTCGGTGTCGCGGAGA 292
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 836 AGTTAGATGGTTTACAGCTATTCCTGTCGACGATGATGCGATCAGCTATTTAGATGG 895
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 AACCGACGCGCTGGAGCGCATTCATCGGATCGGTATCGGCTATTCCTGCGCATGG 232
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 896 GTGATGTTTAAAGTTTAAATGAAAAAGCGCAACAGATGGATCAAGAAAAAGCAAAAG 955
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 GCGAGCTACTGCTCTATCGAAGATATCGAAGCAAAAGTTGACCGCGCACAGGCTGAGA 172
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 956 ATTATAGAG---AAAAAGATCGGTGAGTCACTCGTTTACTTTAGATGATTTTATAGAACAC 1012
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 AGCTGCGACCAACTGAAGAAAGCGCGGTTTCGACCTGAACGACTTCCTGGCAACAGC 112
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1013 TTGATCAGTGAAGAACTAGGACCACTGGATGATATATGAAATGATTCAGAAATGATTCAGGATGA 1072
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 TCAACAGATGAAGAACATGGCGGTATGGCCAGCTCTGATGGGCAATTCACGGGCAITGG 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1073 ATAAATGAAGGGCTAGATGAAGCTTAATATGATGAGTGAAG 1113
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 GGCAGATTCGCGACAACGTTAAATCGCAGATGGATGACAAG 11
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BQ805434 656 bp mRNA linear EST 31-JUL-2002
LOCUS WHE3566_H12_P24ZS Wheat developing grains cDNA library Triticum
DEFINITION aestivum CDNA clone WHE3566_H12_P24, mRNA sequence.
ACCESSION BQ805434
VERSION BQ805434.1 GI:22029643
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 656)
Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
Location/Qualifiers
    1..656
        /organism="Triticum aestivum"
        /cultivar="Butte 86"
        /db_xref="taxon:4565"
        /clone="WHE3566_H12_P24"
        /clone_lib="Wheat developing grains cDNA library"
        /tissue_type="whole grains"
        /dev_stage="3-44 days post anthesis seed"
        /lab_host="E. coli SOLR"
        /note="Vector: Lambda ZAP II, excised phagemid; Site_1:
        EcoRI; Plants were grown under six following different
        environmental regimes in greenhouse, Environment 1)
        24OC/17OC day/night, well-watered, with post-anthesis
        fertilizer, Environment 2) 24OC/17OC day/night,
        well-watered, without post-anthesis fertilizer,

```

Environment 3) 37oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37oC/17oC day/night plus drought, with post-anthesis fertilizer, Environment 6) 37oC/17oC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the 7J Close lab (Chin, close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

BASE COUNT 200 a 115 c 179 g 152 t
 ORIGIN
 Query Match 16.3%; Score 223; DB 14; Length 656;
 Best Local Similarity 58.8%; Pred. No. 7.9e-38;
 Matches 385; Conservative 0; Mismatches 270; Indels 0; Gaps 0;

QY 419 TTTATCGTCCAGCGGATTAATCAATTAACAACAGTAGGGAACAAATGATATTCCTG 478
 DB 1 TTTACAGGCGTCTGCCATGATCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 60
 QY 479 TATACAGTAGGAGAGATCAAGTAAACCCACAAATGTAATGTAATGTAATGTAATG 538
 DB 61 TTTACTCAAGGAGACCGCGCCAAACCTGCAGAAATAACCAAGATGCGGTGAAGAGG 120
 QY 539 CTAACAGACATTTAGATTTGTAATCATGATACAGAGGTCATACACATCGATG 598
 DB 121 CGAAAGAAATATATGATGCGATCGTAGTGATACGTCGCGAGCTGCGATGATGATA 180
 QY 599 AAGCATGTGATGACGAATTAAGAGAGTAAAGAAATGCTAAACCAACGAAATATGT 658
 DB 181 AAACAATGATGTTGTAATGAAAGATTAAGAGGCGATTAATCTCTACAGAGATTTTC 240
 QY 659 TAGTTGTCATGATGACGCGGTCAGATGCTGTCATGTTGCGAGATCTTTTGACGATC 718
 DB 241 TTGTTGTTGATGCCATGCTGCGGCGAGGAGCTGCGAGCATTTGGTCAACAACCTTCAATATG 300
 QY 719 AACTTGATGTCAGAGGTGTTACCTTAACATAATAGATGTTGATACACGTTGCTGCGAG 778
 DB 301 AAATTGGTATATCCGTTGCTATATGACTAAATGGATGGTGATCCAGGCGTGAGCGAG 360
 QY 779 CTTATATCTATTCGTCGCTGACACAAACCAATTAATTTGTTGTTGTTGTTGTTGTTGTT 838
 DB 361 CAGTTAGTGTGAAGAGGTGCTGGAAGCCCATCAAGTTTGTGGCGAGGAGAGCGAA 420
 QY 839 TAGATGTTTAGAGTATTCACCTGAAAGCTGATGCGATCACTGATTTATTTAGTATGGGTG 898
 DB 421 TGGAGACCTTGAGCTTTCTATCCCGACCGCATGGCACACAGTGTGTTGGGAATGGGAG 480
 QY 899 ATGTGTTAGTTTAAATGAAAGCGCACACAGATGTGGATCAGAAAGAAAGCAAGATTT 958
 DB 481 ATGCTCTTCATTTCTTGAAGAGACCAAGAGTGTTCGCCAAGAGGATACCATGGAAC 540
 QY 959 TAGAGAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTATAGACACATCTGATC 1018
 DB 541 TGCAGAGAGAGATCATGAGTGCAGAAATTTGACTTCAATGATTTTAAAGCAGACAAA 600
 QY 1019 AGGTGAAATCTAGGACCATGATGATATTAATGAAATGATTTCCAGGTATGAA 1073
 DB 601 ATGTTGCGAAATGGATCCATGAGCGGTGATCGGAATGATTTCCAGGATGAA 655

RESULT 6
 AW648731 593 bp mRNA linear EST 18-MAY-2001
 LOCUS
 DEFINITION EST327101 tomato germinating seedlings, TAMU Lycopersicon
 esculentum cDNA clone CLEI5G22 5', mRNA sequence.
 ACCESSION AW648731
 VERSION AW648731.1 GI:7409885
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 593)
 AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Roit,I.E.,
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
 Neriman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
 ,S.D.
 TITLE Generation of ESTs from germinating tomato seed
 JOURNAL Unpublished (2000)
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 FEATURES
 Location/Qualifiers
 1..593
 /organism="Lycopersicon esculentum"
 /cultivar="M4496"
 /db_xref="taxon:4081"
 /clone="CLEI5G22"
 /clone_lib="tomato germinating seedlings, TAMU"
 /tissue_type="whole seedlings"
 /dev_stage="7 days post imbibition"
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; 7 days post imbibition on water-agar mixed stage
 whole germinating seedlings from seed coat emergence up
 to two centimeters in seeds not showing obvious signs of
 germination were discarded."
 BASE COUNT 191 a 97 c 159 g 146 t
 ORIGIN
 Query Match 16.2%; Score 221.2; DB 10; Length 593;
 Best Local Similarity 61.8%; Pred. No. 1.9e-37;
 Matches 352; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 442 CAATTACAAACAGTAGGGAACAAATTCATATTCCTGTATACAGTAGGAGAGATCAAGTA 501
 DB 2 CAACCTGTTATTTGGGTAACAGCTTGATGTACCTGTTATGTCAGCAGGAGACAGATGA 61
 QY 502 AAGCCACAAATAATGTAATCAATTAATAACATGCTAAAGAGAACATTTAGACTTT 561
 DB 62 AAACCTGCGAAATAGCCGACAGAGGATTAACAGAGGCCCAAAAGAAAGATGTAGATGA 121
 QY 562 GTAATCATATGATACAGCAGCTGATTAACACATCGATGAAGCATTTGATGAACGAATTAATA 621
 DB 122 GTCATATGATGATCAGCTGGAAGACTTCAGATAGATAAACTATGATGATGAATTAATA 181
 QY 622 GAAGTAAAGAAATGCTAAACCAACGAAATATGATGTTAGTTGTCGATTAATCAACGGGT 681
 DB 182 GACGTGAAGAGGGTACTGAACCCACAGAGGTTTACTTTGTTGTTGTTGTTGTTGTTGTTG 241
 QY 682 CAGATGCTGTCATGTTGTCAGAACTTTTCACCATCACTGATGTCACAGGTGTTACC 741
 DB 242 CAAGAAGCTCGAGCTTTGGTTCACACATTCATCTCGAAATTTGGAATTTACTGGTGCATC 301
 QY 742 TTAATAAATTTAGATGTTGATACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 801
 DB 302 TTGACGAGAGCTAGATGAGATCTAGGGGTGGAGCAGCTTTAAGTGTCAAGGAGGTATCA 361


```

/clone="CLEG63C14"
/tissue_lib="tomato breaker fruit"
/dev_stage="breaker"
/lab_host="SOLR"
/notes="Vector: pBluescriptSKmCuadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT      222 a 119 c 165 g 171 t
ORIGIN
Query Match      15.4%; Score 210.4; DB 13; Length 677;
Best Local Similarity 62.1%; Pred. No. 4.1e-35;
Matches 348; Conservative 0; Mismatches 211; Indels 1; Gaps 1;

QY 453 AGTAGGAAACAATTCGATTCCTGTATACAGTGAAGGAGATCAAGTAAAGCCACACAA 512
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 AGTGGGGAACAGGTTGATGTACCTGTTATGCCAGGACAGATGATAAACCCTGCAGA 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 513 AATTGTACTAATCATTAACATGCTAAAGAGACATTTAGACTTTGTAAATCTGA 572
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 AATAGCCGACAGATTCAGAGGCCAAGAAGAGATGTAGTGTAGTCATATGGA 210
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 573 TACAGCAGTCGATTACACATCGATGAAGCATTTGATGAACGAATTAAGAGAGTAAAGA 632
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TACAGCTGGAAGACTTCAGATAGATAAACTATGATGATGAATTAAGAGAGTGAACG 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 633 AATTGCTAAACCAACGAATTAATGTAGTGTGATTCATATGACGGGTCAAGATGCTGT 692
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 GGTACTGAACCCACAGAGGTTTACTGTTGTGATGCAATGACTGSCCAAGAGCTGC 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 693 CAATGTTCGAGAATCTTTTACGATCACTTGATGTCACAGGTGTTACCTTAACATAAT 752
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 AGCTTTGTCACACATTCATTCGAATTCGAATTTACTGTTGCCATCTTCAGAGCT 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 753 AGATGGTATACACGTCGGTGCAGCTTTATCTATTTCGTTGCGTGACACAAAACCAAT 812
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AGATGGAGATCTTAGGGGTGGAGCAGCTTTAAGTGTCAAGAGGATATCAGGAAGCCAA 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 813 TAAATTTTGTGTAGTGAAGATTTAGATGTTTAGAGCTATTCATCTCGAACGAT 872
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 CRACTCTAGAGAGGGGTGACGTATGGAGACCTTGAACCTTTCTATCTGACCGGCT 510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 873 GGCATCACGATTTTAGG-TATGGGTGATGTTTAAAGTTTAAATGAAAAGCCCAACAG 931
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 GGCTGGACGATTTTAGGAATGGGAGATGTTCTATCGTTGTTGAGAAGCCCAAGAA 570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 932 ATGTGATCAAGAAAGCAAGATTTAGAGAAAGATCGCTGAGTCATCTGTTACTT 991
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 GTATCGGTCAGAGAGATGCTGGAAGATTCAGAGAGAGATCATGAGTGCAAAATTTGAT 630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 992 TAGATGATTTTATAGAACAA 1011
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 TCAATGACTTCTCTGAGCAA 650
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
BE432304
LOCUS      605 bp mRNA linear EST 18-MAY-2001
DEFINITION EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLE67M2, mRNA sequence.
ACCESSION BE432304
VERSION    BE432304.1 GI:9430147
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

```

```

Lycopersicon.
1 (bases 1 to 605)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLE67M2"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/notes="Vector: pBluescriptSKmCuadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT      190 a 97 c 159 g 159 t
ORIGIN
Query Match      13.9%; Score 190.6; DB 10; Length 605;
Best Local Similarity 58.6%; Pred. No. 7.3e-31;
Matches 350; Conservative 0; Mismatches 244; Indels 3; Gaps 1;

QY 192 TCTAATGCAATCATTAACACAGGCAACAGTTATTAAATAGTTCAGATGAATTAAC 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 TTTGATTCGAGGTGTAACACAGATCAACCACTAGTATGACTGTCGTCGAGCTGT 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 GAAGTTGATGGTGGGAGAAATACATCATGATTAATGTCATAAATAAACCCACTTGT 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAACATGATGGTGGAGAGGTTCTGAACTGGTTTTTGTCTAAATCTAAGCCCCACGTAAT 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 312 TATGATGGTGGTGGGAGAAATACATCATGATTAATGTCATAAATAAACCCACTTGT 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACTATTCGCGGCTCTACAAGGTTGGGAGACAACTGTTAGTGCAAGTAGCTTTAT- 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 GATGCGTAAATAATACAAACAAAACCTATGTTAGTTGCGACGAGATATTATTCGTCAGC 431
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 --ATCTAAGNAGCAGGGTAAAGATTGCTGCTGATGCTGGAGCGGTACAGACCTGC 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 432 AGCGATAATCAATTAACAACAGTAGGGAACAATAATGATATTCCTGTATACAGTGAAG 491
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TGCTATTGACCAACTTGTATTATTGGGTAACAAGGTTGATGTAACCTGTTTATGACGAGC 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 492 AGATCAAGTAAAGCCACAAATAATGTTAACTAATGCAATTAACATGCTTAAAGAGAACAA 551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 AACAGATGTAACCTGCGAATAATAGCCCGACAGGATTAACAGAGGCCAAGACAGAA 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 552 TTTAGACTTTGTAATCATTGATACAGCAGGTCGATTACATCATGATGAAGCATTTGATGA 611
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TGTAGATTAGTCATAATGGATACAGCTGGAAGACTTCAGATAGATAAAACTATGATGA 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 612 CGAATTAAGAGAGTAAAGAAATTCGTAACCAACAGAAATATGTTAGTGTGATTC 671
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TGAATTAAGACGCTGAAACGGGTACTGAACCCACAGAGGTTTACTTGTGTGGATGC 477
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 672 AATGACGGGTCAAGCTGCTCAATGTTGAGATCTTTTGGCGATCACTGATGTCAC 731
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 AATGACTGGCAGAGAGCTGAGCTTTTGGTCACAACATCAATCTCGAAATGGAATAC 537
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 732 AGGTGTTACCTTAACATAAATTAGATGTTGATACAGCTGGTGGTGCAGCTTTATCTAT 788
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Query Match 13.6%; Score 186.4; DB 9; Length 581;
Best Local Similarity 58.7%; Pred. No. 5.8e-30;
Matches 341; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

```
QY 111 AGCGTATTATTGAGGCTGACGTAAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGATATC 170
Db 3 AGCACTTTTGGAGCCGATAGTAAAGTTTGCACAGTAGTAGAGATTTATTGAGCTGTGAAG 62
QY 171 AGAAGCCGATTAGGTTCCGATTAATGCAATCAATTAACACAGGAGCAACAAGTTATTA 230
Db 63 TGAAGAAGGCTGAGCCACCGATGATGATGCGAGGTCTCGACCTGACGACAGATTTGGTGA 122
QY 231 AATAGTTCAGATTAATTAAGAAAGTTGATGGGTGGAGAAATACACPCGATTAATATGTC 290
Db 123 GGTGTGTGATGATGAACCTGTTACAACTGATGGCGGGAGGTATPCAGATTTGGTGTTC 182
QY 291 AATATAACACCACTACTGCTGTGTTATGATGGTTTCAAGGTGCTGGTAAACACACAC 350
Db 183 AATATCTGGCCAACTATATATTATTGGCAGGTCTGCAAGGTGTTGGAATAACTACTGT 242
QY 351 TGCAGTAAATTAGCATTTATGATGCGTAAATAAATAACACAAAAAACCCTATGTTGC 410
Db 243 TTCTGCGAAGCTGGCTTCT---ATCTCAAAAAAAGCTGGCAAGAGCTGATGCTAGTTGC 299
QY 411 AGCAGATATTATGCTCAGCAGCGATTAATCAATTAACACAGTAGGAGAAACAATTA 470
Db 300 TGCAGATGTTTACAGGCTGCTGCTATTGATCAACTCACTATTCTGGGTAAAGAGTTGG 359
QY 471 TATTCCTGTATACAGTGAAGGATCAAGTAAAGCCACAAATTTGTAACATTAATGCAAT 530
Db 360 TGTACCAAGTTTACTCAGAAAGACTGACGCAAAACCTTCACAAATAGCCAAAACGGTT 419
QY 531 AAAACATGCTAAGAAAGAACATTTAGACTTTGTAATCATTTGATACAGAGTGGATTA 590
Db 420 GAAGAGGCAAAAGTCAATAGGCTGATGATTAATTAATAGTGGACACAGGCTGGAAGACTGA 479
QY 591 CATCGATGAAGCATTTGATGACGAATTAAGAAAGTAAAGAAATTCCTAAACCAAGCA 650
Db 480 GGTAGATTAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
QY 651 AATATGTTAGTTGCTCAATTAATGACGGGTCAAGATGCTG 691
Db 540 AGTTCTCTGTGGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580
```

RESULT 12
AZ529841
LOCUS ENT0084TR Entamoeba histolytica Sheared DNA linear GSS 03-NOV-2000
DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ529841
VERSION AZ529841.1 GI:11083092
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 814)
AUTHORS Loftus B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 16

High quality sequence stop: 811.
Location/Qualifiers
1. .814

source /organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/vector="pHOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

BASE COUNT 342 a 83 c 172 g 217 t

ORIGIN

Query Match 13.4%; Score 183.4; DB 17; Length 814;

Best Local Similarity 54.1%; Pred. No. 2.6e-29;

Matches 373; Conservative 0; Mismatches 316; Indels 0; Gaps 0;

QY 392 AAAAAACCTATGTTAGTTCAGCAGATATTTATCGTCCAGCAGCATATAATCAATTACAAA 451

Db 32 ATAAACCCAGCTGTTGTTGGAGCAGATACATTTAGCAGCAGCATATGACAAATTACAAA 91

QY 452 CAGTAGGGAACAATTTGATTTCCGTATACAGTGAAGGAGATCAAGTAAAGCCACAC 511

Db 92 TGAATGCTAAACGACGAGGAGTACCATTTTGGGAATAAAGAAAGAGTGTCCAGTAA 151

QY 512 AATTTGTAACTAATGCAATTAACATGCTAAAGAGAACATTTAGACTTTGTAATCATTT 571

Db 152 AAGTTGCTAGTGAAGAGTTCGTACCTTTAGNAGGAAATGATATTAATCTTGTGTG 211

QY 572 ATACAGCAGGTGCGATTACATCGATGAAGCAATTTGATGAACGAATTAAGAAAGTAAAG 631

Db 212 ATACATCAGGTAGACATAAACAAGATAAAGAGTTATTTAAAGAAATGCAAAAGTTCGTG 271

QY 632 AATTTGCTAAACAAACGAAATTTAGTTGCTGATTCATATGACGGGTCAAGATGCTG 691

Db 272 ATGCAATTAACCCAGATTCATATTTTGTAAATGGATGGAGCTTTTGGACAGCAGCAT 331

QY 692 TCAATGTTTCGAGAACTCTTTTACGATCAACTGATGTCACAGGTGTACCTTAACTAAAT 751

Db 332 TTGGACAAACAAAGCATTTTAAAGATGCGAGTTGAAGTAGGAAGTGTATTATAACAAAT 391

QY 752 TAGATGCGTATACACAGTGGTGGTGGAGCTTTATCTATTCTGCTGGTGCACACAAACCAA 811

Db 392 TAGATGCGACATCAAAATGGAGGAGGAGCAATTAAGTCTGTAGCAGCAACAAAGTCCAA 451

QY 812 TTAATAATTTGTTGTTATGAGTGAAGATTTAGATGTTTATGAGCTATTCCTCTGAACTGA 871

Db 452 TTATATTTATTTGGAACAGGAGAAAGTCAATGAATAGAGATTTGTGACAGAAATTT 511

QY 872 TGCGATCAGTATTTTATAGGTATGGGTGATGTTTAAAGTTTAAATGAAAAGCGCAACAG 931

Db 512 TTGTAAGAAAAATTTACTTTGGAAATGGGAGATTTAAAGGAATAGCAAAATAGCTAAAGACT 571

QY 932 ATGTGATCAGAAAAAGCAAAAGATTTAGACAAAAGATGCGTGGATCATCGTTTACT 991

Db 572 TTGAGAGAAATGCTGGAATATAAAACATTTGGTAAACACATCTTCAAGSAGAGCTTAACAG 631

QY 992 TAGATGATTTTTTGAACAACACTTGAATCAGGTGAAAAATCTAGGACCATGGATATTA 1051

Db 632 TAGAGATTTGGAAGACAAATTTCAATTTACAAAAAATGGGACATTTAGGAATATAA 691

QY 1052 TGAATAATGATTCAGGTATGAATATAATG 1080

```

Db 692 TGCATGATGATTGGAATTAATCATCAATG 720
                                     linear      mRNA      HTc 25-MAY-2002

RESULT 13
LOCUS      AY109321
DEFINITION Zea mays P00070230 mRNA sequence.
ACCESSION  AY109321
VERSION    AY109321.1 GI:21212876
KEYWORDS   HTc.
SOURCE     Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 1200)
AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,I.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
JOURNAL   Unpublished (2002)
AUTHORS   Coe,E.C.
REFERENCE  2 (bases 1 to 1200)
JOURNAL   Direct Submission
AUTHORS   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
JOURNAL   Location/Qualifiers
FEATURES   source
            1..1200
            /organism="Zea mays"
            /db_xref="MaizeDB:633818"
            /db_xref="taxon:4577"
            /clone="P00070230"
            /clone.lib="Maize Mapping Project/DuPont Consensus
            Library"
            /note="this sequence is part of a project of EST
            assemblies resulting from the application of public
            contigs to seed DuPont contigs; this resource was
            assembled by DuPont as part of a collaboration for the
            overgo addressing of BACs in conjunction with the Maize
            Mapping Project"
BASE COUNT 387 a 216 c 317 g 279 t 1 others
ORIGIN

Query Match 13.1%; Score 179; DB 11; Length 1200;
Best Local Similarity 55.5%; Pred. No. 2.3e-28;
Matches 391; Conservative 0; Mismatches 305; Indels 9; Gaps 2;

QY 622 GAAGTAAGAATAATGCTTAACCAACGAAATTAATTTAGTTGTCGATCAATGACGGGT 681
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 11 GAAGTGAAGAAGCGCTGTTAACTCCTACAGAAGTTCTGCTTGTGATGCCATGACTGGC 70
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 682 CAAGATGCTGTCATGTTGCAGATCTTTTGACGATCACTTGATGTCACAGGTGTACC 741
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 71 CAAGAGCAGCAGCAGCTAGTACACCTTCATATTCAGATTGGTATACGTGCGTCAATA 130
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 742 TTAATAAATAGATGGTATACACGTTGGTGGTGCAGCTTTATCTPATTCGTTCCGTTGACA 801
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 131 CTGACTAAATGGATGGTGACTCCAGGGCGGAGCGGCACCTAAGGTTTAAAGAGGTTCT 190
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 802 CAAAACCAATTAATTTGTTGGTATGAGTGAAAGCTTAGATGTTAGAGCTTATCCAT 861
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 191 GGGAGGCCCATATAAATTTGTTGGCGTGGGAGCGGATTAGAGACCTTGAGCTTTTCTAC 250
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 862 CCTGAAGTATGTCATCAGTATTTAGTGTATGGTGTGATGTCGTTAAGTTTAATGAAAA 921
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 251 CCTGATCGATGCCACAGGAGTCTCGGAATGGAGAGTGGTGTCTATTTGTGCAAAAG 310
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 922 GCGCAACAAGATGATGATCAAGAAAAAGCAAAAGATTTAGAAAAAGATCGGTGAGTCA 981
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
Db 311 ACACAAGAAGTATGAGGCAAGAGAGCTATAGATTACAGAAGAAGATCATGCTGCA 370
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||
QY 982 TCGTTTACTTAGATGATTTTTTAGACAACCTTGATCAGGTGAAAAATCTTAGACCACGTG 1041
      ||||| || | ||||| || | ||||| || | ||||| || | ||||| || | ||||| ||

```

```

Db 371 AATTTCAGCTTCAACGATTCTCTTAAACAATCTCAAAACGTTTGCAGAAATGGTTCCATG 430
QY 1042 GATGATATTATGAATGATTCACAGCTATGAATAAATGAAAGGCTAGATAACCTTAAT 1101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 431 AGCCGCAATTATTGGATGATGCCAGGCATGAACAAGATAACTCCCGCACAAATCCGAGA - 489
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1102 ATGAGTGAAGAAGCAATGATCATATTAAAGCGATTATCCAGTCAATGACCGCGGCTGAA 1161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 490 --ACTGAGAAAGACTTGCCATTCTGCGGAGTCATGATCAATGCCATGACTGCTGAGGAA 547
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1162 AGAAACAATCCAGACACATGATGATATATCATGTAATAAAGCGTATT-----GCTAAGGG 1215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 548 AGGAGAAGCCAGAGTACTGCTGCTGAATCACTGAGAGGAGGATAAGAGTGCCTGAGGAG 607
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1216 TCTGCTGCTTCATTACAAGAACTCAATCGTTTGTATGAAACAATTTAAGCATATCAAGAAA 1275
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 608 TCTGGAAGACTGAACAAGAGTGAGTCAATTTGTTGCCAGCTTTTCCAAATCGTGCT 667
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1276 ATGATGAACAATTCAGTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 668 CAGATGCAGAGTGTGATGGGTATGGTCARAGGACAGAAGCAATA 712
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 14
LOCUS      BQ996389
DEFINITION OGG12L07.yg.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
ACCESSION  BQ996389
VERSION    BQ996389.1 GI:22430785
KEYWORDS   EST.
SOURCE     Lactuca sativa.
ORGANISM   Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
REFERENCE  1 (bases 1 to 677)
AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,B., Chevalier,P., Ziegler,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://comgenomics.ucdavis.edu/
            Unpublished (2002)
            Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig OG_CA_Contig5305, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: OGG12 row: L column: 07.
            Location/Qualifiers
            1..677
            /organism="Lactuca sativa"
            /cultivar="L.serriola"
            /db_xref="taxon:4236"
            /clone="OGG12L07"
            /clone.lib="OG_EFGHJ lettuce serriola"
            /lab_host="E.coli"
            /note="vector: pBRCNDSfiAB. The library was constructed
            from 10 different sources of RNA from a single genotype.
            Separate cDNAs were generated using primers that
            incorporated unique 5' and 3' tags to distinguish each
            source of RNA. cDNAs were then pooled, size-fractionated,
            directionally cloned into a custom medium-copy vector and
            transformations made with four size classes to minimize
            size bias. Details of each source of RNA and library
            construction can be obtained at http://cgpdb.ucdavis.edu/
            TAG_L1B-OG_EFGHJ lettuce serriola

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 00:20:56 ; Search time 3139 seconds
(without alignments)
4218.470 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFEGSELQATMQKMRGK.....GKKGRNQNMKGNNLFF 455

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09943108/runat_21022003_151630_1801/app.query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -XINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISN=45
-DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -THR SORT=HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943108 -CGN_1_1_3745 -runat_21022003_151630_1801 -NCP0=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_yi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rtd.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2274	100.0	1368	6	E36050 Signal reco
2	2274	100.0	303750	1	AP003133 Staphyloc
3	2274	100.0	347235	1	AP003361 Staphyloc
4	2261	99.4	295350	1	AP004826 Staphyloc
5	1719.5	75.6	2256	6	AX416333 Sequence
6	1717.5	75.5	347050	1	AL591981 Listeria
7	1708.5	75.1	313450	1	AL596170 Listeria
8	1708.5	75.1	319630	6	AX413016 Sequence
9	1708.5	75.1	349980	6	AX417046 Sequence
10	1647	72.4	2739	1	DI4356 Bacillus su
11	1647	72.4	208780	1	BSUB0009
12	1584	69.7	303249	1	AP001515 Bacillus
13	1402	61.7	1136	6	AX433432 Sequence
14	1393	61.3	11024	1	AF007684 Clostridi
15	1382	60.8	296750	1	AF003191 Clostridi
16	1352.5	59.5	14571	1	AE013104 Thermococ
17	1315	57.8	11023	1	AE006391 Lactococ
18	1312	57.7	1605	6	AX194269 Sequence
19	1307	57.5	10462	1	AE008489 Streptoco
20	1306	57.4	1572	6	AX194043 Sequence
21	1306	57.4	7577	6	BD003726 Polynucle
22	1306	57.4	9985	1	AE007428 Streptoco
23	1305	57.4	792	6	E36051 Signal reco
24	1299	57.1	77743	2	SFNEU1910
25	1297	57.0	1569	6	AR081886 Streptoco
26	1297	57.0	1569	6	AR195303 Sequence
27	1297	57.0	1569	6	E35383 Novel ffh
28	1286	56.6	4152	1	U88582 Streptococ
29	1285	56.5	50946	1	AE014153 Streptoco
30	1283	56.4	10173	1	AE010040 Streptoco
31	1279.5	56.3	12187	1	AE008560 Streptoco
32	1238	54.4	11825	1	AE010644 Fusobacte
33	1180.5	51.9	15721	1	AE001802 Thermotog
34	1173.5	51.6	21387	1	AE008821 Salmonell
35	1173.5	51.6	274050	1	AL627276 Salmonell
36	1168	51.4	1641	6	AX122339 Sequence
37	1168	51.4	337200	1	AP005280 Corynebac
38	1168	51.4	349980	6	AX127150 Sequence
39	1164.5	51.2	4586	1	ECTRMD
40	1164.5	51.2	10267	1	AE005491 Escherich
41	1164.5	51.2	11272	1	AE000347 Escherich
42	1164.5	51.2	270365	1	AP002562 Escherich
43	1150.5	50.6	10566	1	AE004142 Vibrio ch
44	1150	50.6	11506	1	AE004793 Pseudomon
45	1139.5	50.1	11004	1	AE013692 Versinia

ALIGNMENTS

RESULT 1

E36050
 LOCUS E36050 1368 bp DNA linear PAT 18-JUN-2001
 DEFINITION Signal recognition particle polypeptide and polynucleotide.
 ACCESSION E36050
 VERSION E36050.1 GI:13022452
 KEYWORDS JP 1999235183-A/1.
 SOURCE unclassified.
 ORGANISM unclassified.

REFERENCE

1. (bases 1 to 1368)

Michael, T.B.

Signal recognition particle polypeptide and polynucleotide

Patent: JP 1999235183-A 1 31-AUG-1999;

SMITHKLINE BEECHAM CORP

OS Unidentified

PN JP 1999235183-A/1

PD 31-AUG-1999

PF 03-SEP-1998 JP 1998289963

PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI

MICHAEL TERENCE BLACK

PC C12N15/09, A61K31/00, A61K31/00, A61K31/00, A61K31/00,

PC A61K31/00, A61K31/00, A61K31/00, A61K31/00, A61K31/70, A61K35/76,

PC A61K39/085,

PC A61K39/395, A61K39/395, A61K48/00, C07K14/31, C07K16/12, C12N1/15,

PC C12N1/19,

PC C12N1/21, C12N5/10, C12N5/10, C12P21/02, C12P1/68, G01N33/15, PC

G01N33/566//

PC C12P21/08, G01N33/577, (C12N15/09, C12R1.445), C12N15/00, C12N5/00,

PC C12N5/00,

PC (C12N15/00, C12R1.445)

CC Strandedness: Single;

CC Topology: Linear;

PH Key

FT source

1..1368

Location/Qualifiers

/organism='Unidentified'.

/organism='unidentified'

/db_xref='taxon:32644'

BASE COUNT 521 a 186 c 288 g 373 t

ORIGIN

Alignment Scores:

Pred. No.: 6,44e-130 Length: 1368
 Score: 2274.00 Matches: 455
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-943-108A-2 (1-455) x E36050 (1-1368)

QY 1 MetAlaPheGluClyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTTGAAGGCTTATCAGACGGCTGCAAGCGACGATGCAAAAATCGTGTGAAG 60
 QY 21 GlyLysLeuThrGluAlaAspPheLysLleMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGTAAACTTACTGAAGCTGATATAAGATATGATGCGTGAAGTAAAGATTAGCGTTATTT 120
 QY 41 GluAlaAspValAsnPhelysValValLysGluPheIleLysThrValSerGluArgAla 60
 Db 121 GAGGCTGACGTAAACCTTAAAGTGGTAAAGATTTATTAAACACAGTATCAGACGGCGA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 Db 181 TTAGGTCCCGATTAATGCAATCATTAACACCGGGCAACAGTAAATTAATAATAGTTCAA 240
 QY 81 AspGlnLeuThrLysLeuMetGlyClyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTAAATATGCAATAAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyValaGlyLysThrThrThrAlaGlyLys 120
 Db 301 CCTACGTTGTTATGATGTTGGTTTACAAGGGCTGGTAAACACAACTCAGGTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTytrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTAGCATTTATGATCGTTAAAAAATACACAAAAAACCTATGTTAGTTGCGACGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATGCTCCAGCAGCATTAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TACAGTGAAGGAGATCAAGTAAACCCACAAATTTGTAACATAATGCAATTAACACATGCT 540
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
 Db 541 AAAGAAAGAACATTTAGACTTTGTATTCATTGATACACAGAGGTGCGATTACACATCGATGA 600
 QY 201 AlaLeuMetAsnGlnLeuLysGlnValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 GCATTGATGAAGCAATTAAGAAGATTAAGAAATTCCTTAACCAACCAAGCAATATGTTA 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 661 GTTCTCGATTCAATGACGGGTCAAGATGCTGTCAATGTTGCGAATCTTTTGACGATCAA 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 CTTGATGTCACAGGTGTTACCTTAACATAAATTAGATGGTGATACACGTGGTGGTCAAGCT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
 Db 781 TTAICTATTCGTCGGTGACACAAAAACCAATTAATTTGTTGTTATGACGAGAAAGTTA 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 GATGGTTTAGAGCTATTCCATCCTGAACGTATGCGCATCAGCTATTTTAGTATGGGTGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 GTGTTAAGTTTAATTGAAGAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAGATTTA 960
 QY 321 GlnLysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
 Db 961 GAGAAAAAGATCGGTGAGTCAATCGTTTACHTTTAGATGATTTTTTAGAACAACTTGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTGAAAAATCTAGGACCATGCGTTCATTACAAGAAAGTCAATCGCTTTGATGAAACAATTT 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
 Db 1081 AAAGGCTAGATAAGCTTAATATGAGTGAAGCAAAATGATCATATTAAGCGCATATC 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 CAGTCATGACCGCGCTGAAGAAACAAATCCAGACACATGGAATGATATCAGTAAAAAG 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATTGCTAAAGGCTCGGTTCGTTCAATTACAAGAAAGTCAATCGCTTTGATGAAACAATTT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
 Db 1261 AACCATATGAGAAATGATGAAACAAATTCACGTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 Db 1321 CGCATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 2

```

AP003133      303750 bp   DNA       linear   BCT 02-JUL-2002
LOCUS
DEFINITION    Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
ACCESSION    AP003133 BA000018
VERSION      AP003133.2 GI:14349175
KEYWORDS
SOURCE       Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315,
              strain:N315) DNA.
ORGANISM     Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE    1
AUTHORS      Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
              Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Iino,T., Ito,T., Kanamori,M.,
              Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
              Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
              Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
              Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
              Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
              Whole genome sequencing of methicillin-resistant Staphylococcus
              aureus
JOURNAL      Lancet 357 (9264), 1225-1240 (2001)
MEDLINE      21311952
PUBMED      11418146
REFERENCE    2 (bases 1 to 303750)
AUTHORS      Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K.
              and Kikuchi,H.
TITLE        Direct Submission
JOURNAL      Submitted (30-JAN-2001) Akio Oguchi, National Institute of
              Technology and Evaluation, Biotechnology Center; 2Chome 49-10
              Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
              (E-mail:oguchienite.go.jp, URL:http://www.bio.nite.go.jp/,
              Tel:81-3-3481-8423, Fax:81-3-3481-8424)
              On Jun 12, 2001 this sequence version replaced gi:13701012.
FEATURES             Location/Qualifiers
     source           1. .303750
                     /organism="Staphylococcus aureus subsp. aureus N315"
                     /strain="N315"
                     /sub_species="aureus N315"
                     /db_xref="taxon:139879"
     gene             202..1155
                     /genes="SA1056"
     CDS              202..1155
                     /gene="SA1056"
                     /note="ORFID:SA1056"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB42308.1"
                     /db_xref="GI:13701013"
                     /translation="MKTLGCLLMLLVAGCSFGNGHKLSSKKSSEKQFTVKES
                     EEKDPDLKYEIEKKMKGIKADSLDKLDPLMEKSFTHSKGICQGWKYEKMGKY
                     ELADYFTDQSGSSIKVDYDFKGGKIKRVIEIHDVDVQVDIWTVPDGGKIGNS
                     NTPVFAEIMTKYDKGKLVYASVEPGSIVIHKDDAKYDDSKLKLKLSQLTKLDHPK
                     VPYSAQIKSFQVPLTSVTFMTHGSKDTKDEVLPALEYTFSPXNYEDKSNFDPKVLN
                     LVMDPLNASSDFGNAHFVLSKYIKESYEVNTASDSSLK"
                     /complement(1318..1521)
                     /gene="SA1057"
                     /note="ORFID:SA1057"
                     /db_xref="GI:13701014"
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="BAB42309.1"
                     /db_xref="GI:13701014"
                     /translation="MQVNVKVIYIILALFLGSGFIHKFYAGKMQGILHLIFCWTGIPH
                     ILAIIASVITVFKPADEQGNVTL"
                     /complement(1820..2308)
                     /gene="SA1058"
                     /note="ORFID:SA1058"
                     /db_xref="GI:13701019"
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="BAB42310.1"
                     /db_xref="GI:13701019"
                     /translation="MAIKKLVPAASHPILTKKAQAVKTFDDSLKRLLOLEDYMYAQEA
                     AGLCAPQINQSLQVAIIDMEMEGLQJLNPKIISQSNETITDLEGSITLPDVIQEVTR
                     SKMIVSYDNGKVELTAHDVARMILHIDQMGIPFTERADRLITDKVEAYFI
                     NP"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB42310.1"
                     /db_xref="GI:13701015"
                     /translation="MAIKKLVPAASHPILTKKAQAVKTFDDSLKRLLOLEDYMYAQEA
                     AGLCAPQINQSLQVAIIDMEMEGLQJLNPKIISQSNETITDLEGSITLPDVIQEVTR
                     SKMIVSYDNGKVELTAHDVARMILHIDQMGIPFTERADRLITDKVEAYFI
                     NP"
                     /codon_start=1
                     /transl_table=11
                     /product="methionyl-tRNA formyltransferase"
                     /protein_id="BAB42311.1"
                     /db_xref="GI:13701016"
                     /translation="MTKIIFMGPPDESTTVLEMLIAEDVIAVVTQDPVGRKRVMT
                     PPVKKVAMKYDLPVYQPEKLSGELEQLQLQDVLIVTAAGQLLPSLALPKLG
                     AINVASLLPKYRGAPLHOAIIDGEGTITIMYKKLDAGNIISQAIKIENDN
                     VGTMDKLSVLGADLLKETLPISIETGNSVPQDQTQATFASNIREDERISWNEPGR
                     QVFNQIRGLSPWPVAVTMDTNLKIYDAELVETKNINERPTIETTKKAILIATNDN
                     EAVAIDKMLAGKKKMLAANYLSGAQNTLVGKKLI"
                     /complement(3155..4540)
                     /gene="SA1060"
                     /note="ORFID:SA1060"
                     /db_xref="GI:13701017"
                     /translation="MSWEKENVSCQLFKCAKHTSRBETMIENVRSLAFDTIQDILN
                     EGAYSNRLINEVSENLNAMDKALFTIYVGTVKRYTLDVLPKPVKTKKAWRQ
                     LLWMSIYOYVLDKVPNHAIINEAVEIAKERGGYHNGVNGILRTMWSDI.PDNEI
                     ADPKRMALIEYSNEKWLIDHWAHYGLEETELIQSLETSTSTVTRANLTRASLDRII
                     EKLQDEGIDKEKDHDLPICLHIGGQPIIHSRSFKDFGVSIODKSSMFVAHIMNVDRHD
                     HVLDCASAPGGKACHIAEVLMPGEGYDASDIHDHDKIDLINFNKIKRLNLNKAQORDA
                     TKPYDKTKILVDAPCSGLGVMRHPEIKYQSKQHIESLVELQLEILENVKNVKI
                     GGEIYVCTTIEQLENENVIYTLKNNKNFEPEFOHETGELVKTQLIMPQDFNSDG
                     FRITKIKRKN"
                     /complement(4543..5637)
                     /gene="SA1061"
                     /note="ORFID:SA1061"
                     /db_xref="GI:13701018"
                     /translation="MITAEKKKKXKLPNFDKQSTYSLSRFDQMNMVLVGGQKFRK
                     QIFEWLQKRVDSIDEMTNLSKDLRLLDNFTVTLTVVKQESKDGIKFLFLOD
                     GYIETVLMRDHYGNSCVCTVOYCGTQFCATGLGKRNLEAGEIYSOVLTVOKYA
                     LDATERYSOIVIMGIGEPENYDNMDFLRVNDNLSLNGAHGHTVSTSGIPIY
                     DPADELIQINFAVSLHAKDEVSRLMPINRAYNVKLEIAIQYQEKINRRVTEYG
                     LFGVNDQLEHARELAHLIKLACHVNLIPVNVHPERNYVKTAKNDIFKFEKELKRLG
                     INATIRREGSDIDAACQLRAKERQVETR"
                     /complement(5644..6387)
                     /gene="SA1062"
                     /note="ORFID:SA1062"
                     /db_xref="GI:13701019"
                     /translation="MLEAQFFFTTQGHDRKDNEDAGGIYNOTNOQLLVLCDSGMGHKA
                     GEVASKFVTDLKSFEAENLIEHQENLNNIKDINFOLYHQAENAEYKMGTT"

```


QY 341 ValLysAsnLeuGlyProLeuAspAspLeuMetLysMetIleProGlyMetAsnLysMet 360
 Db 257750 GTGAAAATCTAGGACCATCGGATGATATTATGAAAATGATTCACGATGATGAAAATG 257809

QY 361 LysGlyLeuAspLysLeuAsnMetSerGlyLysGlnIleAspHisIleLysAlaIleIle 380
 Db 257810 AAAGGCTAGNATAGCTTAATATGATGAAAGCAAAATGATCATATTAACCGATTATC 257869

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 257870 CAGTCAATGACCGCGGTGAAAGAAACAATCCAGACACATTAATGATCATCACTGTAAGG 257929

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 257930 CGTATGCTAAAGGCTGCTGCTGCTCATCAAGAAGTCATCGTTGATGAACAATTT 257989

QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
 Db 257990 AACGATATGAAGAAATGATGAACAATTCATGTCGCGGTAAAGGTAAAGGTAAAG 258049

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 Db 258050 CGCAATCAATGCAAAATATGTTAAAGGTATGATTAACCGTTT 258094

RESULT 4
 AP004826 295350 bp DNA linear BCT 02-JUL-2002
 LOCUS Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,
 DEFINITION strain:MW2, section 5/10.
 ACCESSION AP004826 BA000033
 VERSION
 KEYWORDS AP004826.1 GI:21204263
 SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.
 ORGANISM Staphylococcus aureus subsp. aureus MW2
 Bacteria; Firmicutes; Bacillales; Staphylococcus.
 REFERENCE 1
 AUTHORS Baba.T., Takeuchi.F., Kuroda.M., Yuzawa.H., Aoki.K., Oguchi.A.,
 Nagai.Y., Iwano.A., Asano.K., Naimi.T., Kuroda.H., Cui.L.,
 Yamamoto.K. and Hiramatsu.K.
 TITLE Genome and virulence determinants of high virulence
 community-acquired MRSA
 JOURNAL Lancet 359 (9320), 1819-1827 (2002)
 MEDLINE 22040717
 PUBMED 12043778
 REFERENCE 2 (pages 1 to 295350)
 AUTHORS Aoki.K., Oguchi.A., Nagai.Y., Asano.K., Iwano.N., Baba.T.,
 Kuroda.M., Hiramatsu.K. and Kikuchi.H.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-2002) Akio Oguchi, National Institute of
 Technology and Evaluation, Biotechnology Center; 2Chome 49-10
 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
 (E-mail:oguchienite.go.jp, URL:http://www.bio.nite.go.jp/,
 Tel:81-3-3481-8423, Fax:81-3-3481-8424)
 Location/Qualifiers
 1. 295350
 /organism="Staphylococcus aureus subsp. aureus MW2"
 /strain="MW2"
 /db_xref="taxon:196620"
 301. .1254
 /gene="MW1096"
 301. .1254
 /gene="MW1096"
 /note="ORFID:MW1096"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAB94961.1"
 /db_xref="GI:21204264"
 /translation="MKTLGCLLLMLLVAGCSFGNHLKSSKSEESKQFTVKES
 EEKQPDLEKYEIEKRMKGIKADSLKLDPLMTKSFNSKGIQGWKIDKELMGKY
 ELADYFTKDSGSIKDVDAFFKGGIKRKVIETHDDVKQDYNTVPDGGKIGNS
 NTPVFVAEIMTKYDGLIVASVEPGSVYIHKDDAIKYDDYSLKLSQLTKLDHPKP

VPYVSAQIKSRGVPLTYSVPMTHSGKDTKDEVLPALATFTFSPKNTKSNPDPKVLN
 LVHMFPLNASSDFGNHAFVVLISKYIKYESNTETASDDSLK"
 complement(1417. .1695)
 /gene="MW1097"
 complement(1417. .1695)
 /gene="MW1097"
 /note="ORFID:MW1097"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAB94962.1"
 /db_xref="GI:21204265"
 /translation="MPELKGIIKREANVIISNKGKEMQVKNVIYIILLALFLGSFG
 IHKFVAGNMGGIILHIFCWTGIPHILAIISAVITVFKPADEQGNVTL"
 1919. .2407
 /gene="MW1098"
 1919. .2407
 /gene="MW1098"
 /note="ORFID:MW1098"
 hypothetical protein, similar to polypeptide deformylase"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB94963.1"
 /db_xref="GI:21204266"
 /translation="MAIKLYPASHPIELTKKAQAVIKFDLSKLKLLQDLEDPYAGEA
 AGLCAPQINSLQVAIIDNEMEGLLQLVNPKIISQSNETIIDLEGSTILDFVIGVTR
 SKMVVESYDYNKVELTAHEDVARMILHIIDQNGPIPTERADRLITDKVEAYFI
 ND"
 2400. .3335
 /gene="MW1099"
 2400. .3335
 /gene="MW1099"
 /note="ORFID:MW1099"
 /codon_start=1
 /transl_table=11
 /product="methionyl-tRNA formyltransferase"
 /protein_id="BAB94964.1"
 /db_xref="GI:21204267"
 /translation="MKIIFWGPDPSTVLEMLIAEHVIAVVTQDPFVGRKRVMT
 PPVKKVAMKYDLFVYQPEKLSGELEQLQLDVLIYVTAAGQLIPESLLALPKLG
 AINVHASLLPKYRGGAETHQAIIDGQETGITIMVKKLDAGNIISQAIETENDN
 VGTMDKLSVLGADLLKETPTIIEGTNESVPQDQTATFASNIRREDRIINWPGR
 QVFNQIRGLSPWPVAVTMDDTNLKIYDAELVETINKINEPCTIETTKKAIIVANDN
 BVAYKDMQLAGKRMALAAVLSGAQNTLVGKKLI"
 3332. .4639
 /gene="MW1100"
 3332. .4639
 /gene="MW1100"
 /note="ORFID:MW1100"
 hypothetical protein, similar to RNA-binding Sun protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB94965.1"
 /db_xref="GI:21204268"
 /translation="MIENVRSLAFDTIIDLNEGAYSNLRINEVLSNELNAMDKALF
 TEIVGTVKRKETLDFYLPKPVKTKI4WVQVLLWMSYQVYVLDKVPHEALNAGE
 IAKRGCGYHNGVNGILRPMRMSLDLPFNEIADPKRMAIEISMPKWIIDHWATHYG
 LKRTILOSEFETTTSTVRANLTRASLDIIIEKLQDEYDEKDHDPYCLHIGGQP
 ITHRSFKDGFVSQDSSVFVHINVDHHDHVLDCASAPGGKACHIAEVLHPGGQV
 DASDIHDHIDILINFNLKLRITNKAFOHDTKPYDKTYDKILVDAPCSGLGVNRHK
 PEIKYTSQKHIESLVSLEILELVNKNVKGIGGELIYSTCTIQLNENNVITFLKN
 NKNFEFEPQHPITGELVLTQLINFDQFNSDGFITTKIKKRDN"
 4642. .5736
 /gene="MW1101"
 4642. .5736
 /gene="MW1101"
 /note="ORFID:MW1101"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAB94966.1"
 /db_xref="GI:21204269"
 /translation="MITAEKKKKKELPNFDEKQSIYSILRFDENQNLVLEQGQOKPRAK

Dn	27184	TTATCTATTTCGTTCCGTCACAAAAACCAATTAATTAATTTGGTATGATGAGTAAAAAGTTA	27243
Qy	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Dn	27244	GATGGTTTAGAGCTATTCCATCTGACAGTATGGCATCACGTATTTTAGTATGGGTGAT	27303
Qy	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
Dn	27304	GTGTTAGTTAATTTGAAGAAGCCACACAGAATGGATGGAAGAAAGCAAGAATTTA	27363
Qy	321	GluLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
Dn	27364	GAGAAAAGATCGTGAGTCATCGTTTACTTTAGATGATTTTTTAGAACAACTTGATCAG	27423
Qy	341	ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet	360
Dn	27424	GTGAAAATCTAGGACCACTGATGATATATGAAAATGATTCACAGTATGAATAAATG	27483
Qy	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspPheHisIleLysAlaIleIle	380
Dn	27484	AAAGGCTAGATAACCTTAATATGAGGAAAGCAAAATGATCATATTAAGCGATATTC	27543
Qy	381	GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys	400
Dn	27544	CAGTCAATGACCGCGCTGAAAGAACAAATCCAGACACATGATGATCACGTAAAAAG	27603
Qy	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe	420
Dn	27604	CGTATTGCTAAAGGCTCGTGGCTTCATTACAAGAGTCAATCGTTGATGAACAATTT	27663
Qy	421	AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys	440
Dn	27664	AACGATATGAAGAAATGATGAACAATTCCTCGTGGCGGTAAGGTAAAGGTAAAT	27723
Qy	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Dn	27724	CGCAATCAAAATGCAAAATATGTTAAAGGTATGAAATTTACCGTTT	27768
RESULT 5			
AX416333/c			
LOCUS			
DEFINITION			
Sequence 3324 from Patent WO0228891.			
ACCESSION			
AX416333			
VERSION			
AX416333.1 GI:21448790			
KEYWORDS			
Listeria monocytogenes ATCC 19115.			
SOURCE			
Listeria monocytogenes ATCC 19115			
ORGANISM			
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.			
REFERENCE			
1			
Glaser P. and Kunst F.			
AUTHORS			
Listeria innocua, genome and applications			
TITLE			
Patent: WO 0228891-A 3324 11-APR-2002;			
JOURNAL			
Pasteur Institut (FR)			
FEATURES			
Location/Qualifiers			
1..2256			
/organism="Listeria monocytogenes ATCC 19115"			
/db_xref="taxon:176281"			
BASE COUNT			
533 a 471 c 420 g 832 t			
ORIGIN			
Alignment Scores:			
Pred. No.: 5 36e-96 Length: 2256			
Score: 1719.50 Matches: 331			
Percent Similarity: 86.81% Conservative: 64			
Best Local Similarity: 72.75% Mismatches: 55			
Query Match: 75.62% Indels: 5			
DB: 6 Gaps: 1			
US-09-943-108a-2 (1-455) x AX416333 (1-2256)			
Qy	1	MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Dn	1666	ATGGCATTTGAGGACTACTCTGAGACTCCAGACACATGACAAAATTCGGCGGCAA	1607

Qy	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Dn	1606	GGAAAGTAAACGAGCTGCGTAAAGAAATGATGCGTGAAGTTCGCTTCCTACTT	1547
Qy	41	GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla	60
Dn	1546	GAAGCCGATGTTAACTTAAAGTCGTTAAACAATTTATTAACACAGTAAGCAACGCT	1487
Qy	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Dn	1486	GTGCGCGGAGCGTTATGAAAGGCTTAACCGCGCAACGAAGTTATCAAAATCGTCAA	1427
Qy	81	AspGluLeuThrLysLysMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Dn	1426	GAAGAACTAACAGCTTAATGGCGGAGAAAGCAAAATCGGACAGCGGACCGCCCA	1367
Qy	101	ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Dn	1366	CCAAACGCTGATTGATGTTAGGTTTACAAGGGCTGGTAAACACAGACTTCGSCCAA	1307
Qy	121	LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAspIle	140
Dn	1306	CTTGCTAATTTTATACGCAAAAAATATAACGTAACCACTTACTTGTCCGACGATAIT	1247
Qy	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Dn	1246	TACGACCTCGCGCAATCAACAATTAGAACACACTTGGCAACCAATTAGATATGCGGTA	1187
Qy	161	TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla	180
Dn	1186	TTTTCTTAGAGACCAAGTAAGCCGCTAGAAATCGCCAAACAGCTATTGCTAAAGCA	1127
Qy	181	LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Dn	1126	AAAGAGAACATTTAGATTATGTTATTATGATACAGCGGCTCTTTCATATTAGCAA	1067
Qy	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Dn	1066	ACGTAATGATGATGTTAAACAAGTGAAGAAATTCGAACCCCAACCGAAATCCTGCT	1007
Qy	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln	240
Dn	1006	GTTCGATTCATGACTGGGCAAGATGCTCAACGTGGCGCAAAAGCTTCAACGAACA	947
Qy	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla	260
Dn	946	TTGAAATACAGGTGCTGCTAACCACCAATTTAGACGCTGATACACGCTGGTGGCAGG	887
Qy	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Dn	886	CTTTCGATTCGTTCCGCTACTGGCAACCAATTAATTTATCGGACTCGTGAACAAATG	827
Qy	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Dn	826	GAAGCTCTAGAACTTCCATCCAGATCGTATGGCTTCCAGAATTCGCGCATGGCGAT	767
Qy	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
Dn	766	GTGCTTTCCCTGATTGAAAAAGCGCAACCGATGTAGATACAGAAAAATGAAGCTATG	707
Qy	321	GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln	340
Dn	706	GAACAAAAATGAAGACACAGCATGCGCTAGATGACTTCTTAGAACACAGTTACACAA	647
Qy	341	VallLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Dn	646	GTAACAAATGGGACCACTAGATGACTTAAATGATGCCAGGGGCAACAAATG	587
Qy	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle	380
Dn	586	AAAGGCTCGACACACATGAACGTCGATGATAAACAACACTAGGTCATATTGAAGCAATTATC	527

```

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
:::|||||  |||:::|||||  ::|||  |||||:::|
Db 526 AAGTCATGACCAAAATGAAAACACATCCAGATATCATCAATCCAGTAGAGAAA 467
|
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||:::|||||  |||:::|||||  ::|||  |||||:::|
Db 466 CGGATTCCAGTGGTACGCGACGTCGCGTTCAAGAAATTAACCGTCTCTAAACAATTT 407
|
QY 421 AsnAspMetLysMetMetLysGlnPheThrGlyGlyGlyGlyGlyGlyGlyLys 440
:::|||||  |||:::|||||  |||:::|||||  |||:::|||||  |||:::|||||  |||
Db 406 GGTGAATGAAAAAATGATGAAGCAATGACTGGTGGCGGCAAGGCAAAAAGGCAAA 347
|
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 346 -----AATCCATTGGCAATTCAAAATGCCATTT 317
|
RESULT 6
AL591981/c 347050 bp DNA linear BCT 06-JUN-2002
DEFINITION Listeria monocytogenes strain EGD, complete genome, segment 9/12.
ACCESSION AL591981 AL591824
VERSION AL591981.1 GI:16411141
KEYWORDS
SOURCE Listeria monocytogenes.
ORGANISM Listeria monocytogenes
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
Baquero,F., Bloeker,P., Bloeker,H., Brandt,P., Chakraborty,T.,
Charbit,A., Chetouani,F., Couve,E., de Baruvvar,A., Dehoux,P.,
Dumann,E., Dominguez-Bernal,G., Duchaud,E., Durand,L.,
Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
Jackson,D., Jones,L.M., Kaerst,U., Krefz,J., Kuhn,M., Kunst,F.,
Kurapat,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.,
Nedjari,H., Nordieik,G., Novella,S., de Pablos,B., Perez-diaz,J.C.,
Purcell,R., Remmel,B., Rose,M., Schlueter,T., Simoes,N.,
Tierrez,A., Vazquez-Bollan,J.A., Voss,H., Wehland,J. and Cossart,P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
JOURNAL Science
MEDLINE 21537279
PUBMED 11679669
REFERENCE
2 (bases 1 to 347050)
AUTHORS Glaser,P., Frangeul,L. and Rusniok,C.
TITLES Direct Submission
JOURNAL Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
COMMENT E-mail: pglaser@pasteur.fr
FEATURES
source Location/Qualifiers
1. 347050
/organism="Listeria monocytogenes"
/strain="EGD-e"
/db_xref="taxon:1639"
complement(95..841)
/gene="lmo1688"
complement(95..841)
/gene="lmo1688"
/note="similar to glucose 1-dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAC99766.1"
/db_xref="GI:16411142"
/db_xref="SPTREMBL:O8Y6J6"
/translation="MKVYVTSRRGLGRETALAKEGYDAVNFNRKKAEVQ
QEQIGKRCVIFRANGVDVKEVRFKAVDEEGRDLDFINRAAGVLRPLMELES
HEDWMTNKAALIFAGQEAAKLMQHQSGKXISLSIGSTIRYLENTYGVSKAAE
SLTRYLAVELAPFGIAYNAVYSGGLIETDLNHFNRRELLKDAVSKTPAGRIEPNDL
VNAVILFLASERADMRGOTILVDGGRTILV"
complement(join(845..1942,1948..1953))
/gene="lmo1689"
complement(845..1942)
/gene="lmo1690"
complement(845..1942)

```

```

/gene="lmo1689"
/note="similar to A/G-specific adenine glycosylase"
/codon_start=1
/transl_table=11
/protein_id="CAC99767.1"
/db_xref="GI:16411143"
/db_xref="SPTREMBL:O8Y6J5"
/translation="MRKMLTWDTKITAFQZALVSVFANKRVLPRNTEPVRILW
VSEIMLOQTQVDVIVFYFRFTQPTPMESFYNEADAILKAWEGLYYSVRNLQTA
IKQVMDAPSGSEVFDLTLLSLKRGVPTAGAILSLIAYNOAFNDGNVFRVIRKLE
IGEDLMKASTKIFEEVLYOLDKNPAFNOGLMEIGALVCTPTKPMCLCPLOPC
EAKKGVETNPVKIKKVKMTKELISIVISEDGKIAIEKRPENGLIANWQPTIE
ISKNEDEVAKLOFLHNYGLEVLIDEPTAHIKHVFSLVVKMDIRVAKLOSAIPNEN
WYFATEEMKRLAPPVYOKMQQAKWDFKE"
complement(850..855)
/gene="lmo1688"
complement(850..855)
/gene="lmo1688"
complement(1948..1953)
/gene="lmo1689"
join(2129..3123,3130..3147)
/gene="lmo1690"
2129..2134
/gene="lmo1690"
2143..3123
/gene="lmo1690"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99768.1"
/db_xref="GI:16411144"
/db_xref="SPTREMBL:O8Y6J4"
/translation="MDTGVHVYMGIALGALATVDPVAVGSSQAAGIMPTAILIGSQIP
DIDTVLTKNNADYIRNRHGTSHSLPMLAIWPLLISILYLPPAATFHLIAWPIA
VGLHLPVDFNAVCTQAVRFKXTWVAFGFTFDWFIRGSHVVAALWLLGSPVLPT
FVTLIALIYIAKVFVQRMKHAVONLIPDSEELIISITFIHFQWRVAVTKDHYI
VGRAKRNIISYKFDRLPVPDNETIRSAKKDKNLAAFISFKVYNWREEKLDGTVY
TPTDLYRNGHPFVAVVKLDDDLKIYSSYTGWIFSTEKLYKLAPVSI"
complement(join(3130..3147,3156..3617,3625..3630))
/gene="lmo1691"
3130..3147
terminator
/gene="lmo1690"
complement(3130..3147)
/gene="lmo1691"
complement(3156..3617)
/gene="lmo1691"
/note="similar to deoxyuridine triphosphate
nucleosidohydrolases"
/codon_start=1
/transl_table=11
/protein_id="CAC99769.1"
/db_xref="GI:16411145"
/db_xref="SPTREMBL:O8Y6J3"
/translation="MKVRFVFNASRKFEQIISLPIRGKSGAGIDFFSNFTVTL
VPEKHFIDKSIWOEDVLNIVYRSIGIKKGLICNGTGIIIDSIYSNPGNDGN
IGIAIKFNFSNPFYSIEAGRVAGVQKYLIVADTIDVANSRYGGVGSTGR"
complement(3625..3630)
/gene="lmo1691"
complement(3629..3862)
/gene="lmo1692"
complement(3629..3862)
/gene="lmo1692"
/codon_start=1
/transl_table=11
/protein_id="CAC99770.1"
/db_xref="GI:16411146"
/db_xref="SPTREMBL:O8Y6J2"
/translation="MTYIDELTAELMYKNFHLDEQALWVIELWMSDFESSYAKAG
PYRGPEYAADVVRQIERHGSFTJHOVRKDPNK"
complement(join(3864..4673,4680..4685))
/gene="lmo1693"
complement(3864..4673)
/gene="lmo1693"

```

```

/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99771.1"
/db_xref="GI:16411147"
/db_xref="SPTREMBL:O8Y6J1"
/translation="MKITISVQOKNKERYNIFIDIRYKNSVDEEVLARYOLMKGVKL
TEADIEREKQADMYRKLGNALNFIASHRVRSEKIRIDYLRKQEMPEFAIDELKLAN
MDYINDEFAELTKQIKTLKGPRTIERELVEKGLTREIISRVLSEYSDAELENA
TKQAKIMKRNKNSAKMLQOKIITLDLQKFTSELAKAAATSELDVADERAILQ
KOIETMKNNKRYKPSAKOKITSLMQKGFSDYTIQSYLTENEISFEEEE"
/gene="lmo1692"
/complement(3872..3877)
/gene="lmo1692"
/complement(3872..3877)
/gene="lmo1692"
/complement(4680..4685)
/gene="lmo1693"
/join(4757..5673,5681..5708)
/gene="lmo1694"
/4757..4762
/gene="lmo1694"
/4771..5673
/gene="lmo1694"
/note="similar to CDP-abequose synthase"
/codon_start=1
/transl_table=11
/protein_id="CAC99772.1"
/db_xref="GI:16411148"
/db_xref="SPTREMBL:O8Y6J0"
/translation="MHILITGATGFTGDLVHELEKSDHLYITLRQKLNKRNHYHI
EWNLNDLMDPEVDVPCINLAGLMDKWTYDRKVTNRSRIEATGALLSIVKMK
KSKPLWNAIGAIVTSSKSTIYLDRENTYADNFKGVYSEKTAASADIGIRV
VYARGLVLTNGSGSPVFEKLEPQITGGRFGNGRQWFSWIHVDVVAAILIPDHEQ
INGVNTAPHPVDEKFAERLQKMKHPYKATVPFKIIRFILGERAMTILDSQRAYP
EKLMSNHEFFEFTELQALDDLLD"
/5681..5708
/gene="lmo1694"
/complement(join(5682..8291,8299..8304))
/gene="lmo1695"
/complement(5682..5709)
/gene="lmo1695"
/complement(5694..8291)
/gene="lmo1695"
/note="similar to putative membrane proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99773.1"
/db_xref="GI:16411149"
/db_xref="SPTREMBL:O8Y6I9"
/translation="MKEKLMOAVAFQKNSTVVKIVFTTFVMAFVIFELINATGIDY
PSLKNLSQSPELFTMFIVGLIAYTPMLLYDYVIKLLPKGKSPSHVTSAGTNT
FTNIGSGVIGLSRASFTGKNASHKEILLALSKIALFLVSGLSYICVLSLATLIP
GRADHFVNWPLLAGDLPFLIFTITKWSKSLFVDLPKRELTLLIISLLEWGFAP
GCFATILMGEPVDIPKPLFVIVASVIGIASVPGGVGTFDVMILGSLQVSOE
LALAMLLFRIFYIIPVFWGLLFFQKAGKVNDFLEGFLFLQKVAHRLVIFVY
GSLGILLSSAPNAIVPFLYKIMPENFTSOITIVAFGLLGLLARGTECKTKK
ATIIIVLGCATENTLARYFSMKQALFGLVLLCLPLARNEFYREKLYVYTSKVIID
SIIFIVLCAIVYIGIYNSPNIKHSEIIPYIRIASHELWGVGVFIVAVSVLIIV
IYLTSTKELSGSPFAVYREHLAKWGNVESHMFLDKLLFWAAGEVLEFSYRILIA
DKMVGTEPTGMNKKMAEIEEYMNADRFYRPFVYEVGTMTPIYIHDHGFDTLKL
EEGFVDVQNTMSKKKKGERALMKDLREGYTFEIIIEPPNFHDWTWTRAVSDWLD
PSGIMDFSLFNLPEKAKEDGPTFNAGMAPLANGVESKYAFLEGRAGLVYRSOGFY
GPKGLNFKSKYIVENEQKVFARFKRSSIAFTMLQIMLVCKRKLANSOVLDPLE
EETKPKDSE"
/complement(8299..8304)
/gene="lmo1695"
/complement(8318..9007)
/gene="lmo1696"
/complement(8318..8992)
/gene="lmo1696"
/note="similar to unknown proteins"

```

```

Alignment Scores:
Pred. No.: 2,22e-93 Length: 347050
Score: 1717.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best Local Similarity: 72.75% Mismatches: 55
Query Match: 75.53% Indels: 5
DB: 1 Gaps: 1

US-09-943-108a-2 (1-455) x AL591981 (1-347050)
QY 1 MetalapheduGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 122116 ATGGCATTTGAAGGACTAGCTGGAAGACTCCAGAAACATGAAACAAATTCGGGGTAAA 122057
QY 21 GlyLysLeuThrGluAlaAspIleyleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 122056 GGGAAAGTAACCAAGACTGACGTAAAGAAATGATCGTGAAGTTGCGCTTGCTCTACTA 121997
QY 41 GluAlaAspValAsnPhelYsValLysGluPheIleLysThrValSerGluArgAla 60
Db 121996 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAAACAGTAGTAAGCGAGCT 121937
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 121936 GTCCGGCGGAGCGTTATGAAAGCCTAAACCCGCCACACAGTTATCAAAATCGTTCAA 121877
QY 81 AspGluLeuThrLysLeuMetGlyGluAspThrSerIleAsnMetSerAsnLysPro 100
Db 121876 GAAGAAGTAACAGCTTAATGCGCGGAGAGAAAGAAATCGGAACAGCGGCGCGCG 121817
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 121816 CCACCCGCTGATTGATGATGAGTTTCAAGGGCTGCTTAACACAGACGACITCCGGGCAA 121757
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 121756 CTTGCTAATTTTACGAAAATAATATACCGTAAACCATTTACTTCCACAGCATATT 121697
QY 141 TyrArgProAlaAlaLailaLeAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 121696 TATCGACCTGCCGGAATCAACAATTAAGAACCTTGGCAAGCAATTAGATATGCGGTA 121637
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 121636 TTTTCCCTAGGAGACCAAGTAGCCGCGTCAAGTCCCAACCAAGCTATTCTTAAAGCA 121577
QY 181 LysGluGluHisLeuAspPheValIleleAspThrAlaGlyArgLeuHisleAspGlu 200
Db 121576 AAAGAAGAACAATTTAGATTATGTCTATTATCGATACAGCCGCTGCTTCTCATATTGACGA 121517
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 121516 AGCTTAATGGATGAGTTTAAACAAGTGAAGAAATTCACAGCCACACCGAATCTTACTT 121457
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 121456 GTAGTCGATTCCATGACCGGCGCAAGTGTCTCAACGTGGCGCAAGCTTCAACGAACA 121397
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 121396 TTAGAAATACAGGTGTCGTTATTAAGAACTAGATGTTGATACACGTGTGTGTGCCGCG 121337
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 121336 CTTTCGATTCGTTCCGTCACCTGCTTAACCAATTAATTTATCCGACTGCCGAAAAATG 121277
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 121276 GAAGCCCTTGAACCTTCCATCCAGATCGTATGCTTCCAGAAATTCCTCGCATGGGGAT 121217
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 121216 GTGCTTTCCTGATTGAAAAAGCGCAACCCGATGTAGATACAGAAAAAATGAAGCTATG 121157

```



```

Db 138781 TTAGAAATTACCGCGTGTATTAAACAAATTAGACGGTGATACACGTCGGTGGGCAGCA 138722
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 138721 CTTTCATCCGTTTCAGTCACAGAAACCAATCAATTCGTTGCTACCGGTGAAAAATG 138662
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 138661 GAAGCAATCAAGAAACGTTCCATCCGATCGTATGGCTTCAAGAATTCGCGCATGGGTGAT 138602
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGlnLysAlaLysLeu 320
Db 138601 GTACTTTCTCTATGTGAAGAAGCACAACTGATGTAGATGCAGAAAAATGAAGACTATG 138542
QY 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
Db 138541 GAACAAAAATGAAGACACACGATGACCTTAGATGACTTCTTGACCAANTTGCACAA 138482
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 138481 GTAAACAAATGGGACCACTAGATGAATCTATAAAATGATGCGAGGGCAACAAATG 138422
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 138421 AAGGCTCGCAACATCGAAGTAGATGATAACAACTCGGTCACTCGAAGCGATAAT 138362
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 138361 AAATCCATGACCAAAAACGAAAGAAATATCGGACATCATCATCGGACGAGAGAAA 138302
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 138301 CGAATTCCTGCGTGAAGCGGACGCCAATTCAGAAATCAATCGCTCTCTAAACAAAT 138242
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysLysLys 440
Db 138241 GCTGAATGAAGAAATGATGAAGCAATGACTGGTGAGGAAAGATGAAGAAAGTAAA 138182
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 138181 -----AATCCATTCGGCAATTTCAAAATGCCATT 138152

RESULT 8
AX413016 LOCUS AX413016 319630 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Parent: WO 0228891-A 7 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
Location/Qualifiers
1. .319630
/organism="Listeria innocua"
/db_xref="taxon:1642"
BASE COUNT 105207 a 55428 c 66726 g 92263 t 6 others
ORIGIN

Alignment Scores:
Pred. No.: 7 12e-93 Length: 319630
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

```

US-09-943-108a-2 (1-455) x AX413016 (1-319630)

```

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 187924 ATGGCATTTGAAGACATAGCTGGAGACATCCCAAGAAACAATGAACAAATTCGCGTAAA 187983
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 187984 GGGAAAGTAAACGAAGCTGACGTAAAGAAATGATGCGTGAAGTTCGTCTTGCTTACTA 188043
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 188044 GAAGCCGATGTATTAAGTCGTTAAACAATTAATAAACAAGTAAACAGTAAACGAAGTCA 188103
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 188104 GTTGGCGGGACGTATGAAGAAACCTACACCGGTCAACAAGTAAATTAATTCCTTCA 188163
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 188164 GAAGAATTAACAAGCCTCATGGCGCGGAAGAAAGTAAATCGGAACAGCTGATCGCCG 188223
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrThrLagLys 120
Db 188224 CCAACCGTTATTTATGATGGTAGGTTTACAAGGAGGTGGTAAAAACAACCACTTCAGGAAA 188283
QY 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysProMetLeuValAlaAlaAspIle 140
Db 188284 CTCGCTAATTTATTTACGTAATAAATAATATATCGTAACCTTTACTAGTCGACGACATAT 188343
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 188344 TACCGACCTGCGACGAATCAACAACATTAGAAACACTGTGCAACAATTAGATATGCCAGTA 188403
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 188404 TTTTCTTAGCGCATCAAGTAAGCCCAAGTAGAAATCGGGAACAAGCTATCGCTAAAGCA 188463
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 188464 AAAGAAGAATTTAGATTATGTCATTATCGATACAGCTGCTCTTCATATCGACGAA 188523
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 188524 ACTCTGATGGAGCAATTAACAACAAAGTGAAGAAATCGCTAGCCCACTGAAATTTACTT 188583
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 188584 GTAGTTGATTCATGACTGGCGAAGCGGTAATGGGCCCAAGCTTCAACGAACAA 188643
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 188644 TTAGAAATTCGCGCTGTATTAAACAAATTTAGACGGTGATACAGTGGTGGGCAGCA 188703
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 188704 CTTTCATCCGTTTCAGTCACAGGAAACCAATCAAAATTCGTTGCTACCGGTGAAAAATG 188763
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 188764 GAAGCAATCGAAACGTTCCATCCGATCGTATGGCTTCAAGATTCCTCGGATGGGTGAT 188823
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValaspGlnGlnLysAlaLysLeu 320
Db 188824 GTACTTTCTCTATTGAAAAAGCACAACTGATGTAGATGCAGAAAAAATGAAGACTATG 188883
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 188884 GAACAAAAATGAAGACAAACATGACATGACCTTAGATGACTTCTTGGACCAATTCGAACA 188943
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 188944 GTAAACAAATGGGACCACTAGATGAATCTATAAAATGATGCCAGGGGCAACAAATG 189003

```

```

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189004 AAAGCCTCGACAAAGATGCAAGTAGATATAACAAGTCCGTCACATCGAAGCCGTAATT 189063

QY 361 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189064 AAATCCATGACCAAAACGAAAGAGATAATCCGACATCATCAATCGCAGCAGAAGAAA 189123

QY 401 ArgIleAlaLysGlySerClyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189124 CGAATTCCTCGTGGAGCGGAGCCCAATTCAGAAATCAATCGCTCTTAAGCAATTT 189183

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189184 GCTGAATGAAAAATGATGAGCAATGACTGCTGGAGGAAAAAGGTAAAGGTAAA 189243

QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189244 -----AATCCATTCGGCAATTTCAAAATGCCATTT 189273

RESULT 9
LOCUS AX4117046 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4037 from Patent WO0228891.
ACCESSION AX4117046
VERSION AX4117046.1 GI:21449656
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Glaser,P. and Kunst,F.
TITLE Listeria innocua genome and applications
JOURNAL Patent: WO 0228891-A 4037 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2038, original length: 3.011.208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101055 a 72969 c 60688 g 115268 t
ORIGIN

Alignment Scores:
Pred. No.: 7,89e-93 Length: 349980
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX4117046 (1-349980)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129501 ATGGCAATTTGAGAGACTACTGGAAGACTCCAGAAACATGACAAATTCGGCGGTAAA 129442

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129441 GGGAAAGTAAACGAAGTCACCTAAAGAAATGATGGGTGAAGTTCGCTTGCTTACTA 129382

QY 41 GluAlaAspValAsnPheLysValVallysGluPheIleLysThrValSerGluArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129381 GAAGCGATGTAACTTAAGTCGTTAAACATTTATTAACACAGTAAGCGAACGTGCA 129322

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129321 GTTGGCCGCGAGCTATGAAGAGCCTACACCCGTCACACAGTATTAAATGTTCAA 129262

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129261 GAAGAATTAACAAGCCTCATGTGGCGGCGAAGAAATATAATCGAACACGCTGATCGCCG 129202

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129201 CCACCGCTTATTATGATGTAGTTTACAGAGCTGGTAAACACACCTTCAGGAAAA 129142

QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129141 CTCGCTAATTATTACGTAAAAAATATAATCGTAACCTTTACTAGTCGACGAGATATT 129082

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129081 TACCGACTGCGAGCAATCAACAATAGAACACTTGGCAACAATTAGATATGCCAGTA 129022

QY 161 TyrSerGluGlyAspGlnVallysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129021 TTTTCTTAGCGGATCAAGTAAGCCAGTAGAATCCGAAACACAGCTATCGCTAAAGCA 128962

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128961 AAAGAAGACATTTAGATTATGCTATTATCATGATACAGCTGGTGGCTTCATATCGAGAA 128902

QY 201 AlaLeuMetAsnGlnLeuLysGluVallysGluIleAlaLysProAsnGluIleMetLeu 220
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128901 ACTGTGATGGCAGAAATTAACAAGTGAAGAAATCGTCACGCCAACTGAAATTTTACTT 128842

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128841 GTAGTGATTCAATGACTGGGCAAGACGCAATAATGTGCCCAAGCTTCAAGCAACA 128782

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128781 TTAGAANTACCGCGCTGTATTACAAATACAGCGTGATACAGTGGTGGGCGACGA 128722

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128721 CTTTCCATCGGTTCACTACAGAGAAACCAATCAATTCGTTGCTACCGGTCGAAAAAATG 128662

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetClyAsp 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128661 GAAGCAATCGAAACGTTCCATCCGATCGTATGCTTCAAGAATTCGCGCATGGGTGAT 128602

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128601 GTACTTCTCTTATGAAAAACACAACTGATGATGATGCAAAAAAATGAAGCTATG 128542

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128541 GAACAAAAAATGAAGACACACGATGACCTTAGCTGACCTCTTGACCAATTCGACAAA 128482

QY 341 VallysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128481 GTAAAACAAAATGGGACCACTAGATGACTACTAAAAATGATCCAGGCGCAACAAAATG 128422

QY 361 LysGlyLeuAspLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128421 AAAGCCTCGACACATGCAAGTAGATGATATAACAACCTCGGTACATCGAAGCGATATT 128362

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128361 AAATCCATGACCAAAACGAAAGAGATAATCCGACATCATCAATCGGAGCAGAAGAAA 128302

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128301 CGAATTCGTGTGGAAGCGAGCCCAATTCAGAAATCAATCGCTCTTAAGCAATTT 128242

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db 128241 GCTGAATGAAAAAATGATGAAGCAAAATCACTGGTGGAGAAAAGGTAGAAGGTAAA 128182

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455

Db 128181 -----ATGCATTCGGCGCAATTCCTCAAAATGCCAATT 128152

RESULT 10

D14356

LOCUS

DEFINITION

D14356 2739 bp DNA linear BCT 03-FEB-1999

Bacillus subtilis orf1, ffh, rps genes for ORF1, ffh and 30S

ribosomal protein S16, complete cds.

ACCESSION

D14356

VERSION

GI:439700

KEYWORDS

orf1; 30S ribosomal protein S16; ffh.

SOURCE

Bacillus subtilis (strain:168) DNA.

ORGANISM

Bacillus subtilis

REFERENCE

1 (bases 1 to 2739)

Honda,K., Nakamura,K., Nishiguchi,M. and Yamane,K.

Cloning and characterization of a Bacillus subtilis gene encoding a

homolog of the 54-kilodalton subunit of mammalian signal

recognition particle and Escherichia coli Ffh

J. Bacteriol. 175 (15), 4885-4894 (1993)

93328695

2 (bases 1 to 2739)

Yamane,K.

Direct Submission

Submitted (03-FEB-1993) Kunio Yamane, Inst. Biological Sciences,

Tsukuba University; Tsukuba, Ibaraki 305, Japan (Tel:0298-53-6419,

Fax:0298-53-6006)

FEATURES

source

location/Qualifiers

1..2739

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

292..297

/evidence=experimental

315..320

/evidence=experimental

365..697

/gene="orf1"

365..697

/gene="orf1"

/codon_start=1

/transl_table=11

/product="ORF1"

/protein_id="BAA22221.1"

/db_xref="GI:2424968"

/translation="MSLEKTRRMNLYDFYQSLLTSKQSYMSLYLLDDFSLGELIAEE

YEVSQAVDNIKRTTEAMLEQIEBKLLLLKKFQERKENFKELASGSKEEELITAL

IEALEKLD"

711..2051

/gene="ffh"

711..2051

/gene="ffh"

/codon_start=1

/transl_table=11

/product="ffh"

/protein_id="BAA21691.1"

/db_xref="GI:2309080"

/translation="MAFESGLADRLQQTISKIRGKGVSEQDVKEMREVRLLAEADV

NFKVVDYFKYSERAVGQVYMKSLTPGQVIVKQVEELTELMGSEESKIAVAKRPPT

VIMVGLQGAGVTTSGLKLANLKKHNRKPMIAVYRPAATKOLETIGKOLDMPV

FSIGDQVSPVETAKQAEKAEHYDVILDTAGRLHDHMLDELINVEIANPEI

FLYVDSWQDVAIVKAFSNEQLGLTGVLTLDGDTRGGAALSIRAVNTPIKFGAL

GEKLDLEFFHPERNAIILMGDVLTLEKAQASVDEDKAKELEQKMTMSTLDDF

LEQIGVRNMGPLDLEQMFGAGKMKGLNTQVDEKOLNHEVAIIKSMVLEKEQPD

IINASRRKRIANGSGTSVQEVNRLKQFDEMKMKMQTNNSKGGKFKLPTM"

2091..2096

-35_signal

2117..2122

-10_signal

2157..2429

/gene="rpsp"

2157..2429

CDS

/gene="rpsp"

/codon_start=1

/transl_table=11

/product="30S ribosomal protein S16"

/protein_id="BAA21692.1"

/db_xref="GI:2309081"

/translation="MAVKIRLKRMGAKKSPFYRIVVADSRSPDRGRLETVGTYNPVA

KPAEVIKIDELALKWLQTKGAPSDTVNLFSSQIGIMEKFNHAKQK"

BASE COUNT 887 a 527 c 648 g 677 t

ORIGIN

Alignment Scores:

Pred. No.: 1,69e-91 Length: 2739

Score: 1647.00 Matches: 319

Percent Similarity: 84.62% Conservative: 56

Best local Similarity: 70.11% Mismatches: 60

Query Match: 72.43% Indels: 10

DB: 1 Gaps: 1

US-09-943-108A-2 (1-455) x D14356 (1-2739)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20

Db 711 ATGGCATTTGAAGGATTAGCCGACCTGCAGCAGACGATTCTTAAATCCGCGAAAA 770

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

Db 771 GGGAAAGTCAGCGAACAGATGTAAGAGATGATGCGTGAGGTCCTGTGCGCTGCTT 830

QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60

Db 831 GAGGCTGACGTTAACTTTAAAGTAGTCAAGGATTTGTCAAAAAGTAAGTGAACGCGCT 890

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80

Db 891 GTAGGCCAAGACGTCATGAAAGCTGACGCCGCCAGCAGGAGTTCATTAAAGTTGTTCAA 950

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100

Db 951 GAGGAACAGTACGTAGCTGATGGCGGCGGAGAGACAAATCCGCCGCAAAAGGCGG 1010

QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120

Db 1011 CCGACTGTTATTATGATGTCGCTCCCAAGGTGCGGTAAACGACACCAACGCGSTAAG 1070

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

Db 1071 CTTGCGAATCTGCTGCGCAAAAAGCATAATCGCAACCGATGCTGTTGTCGCGATATT 1130

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160

Db 1131 TACGCCCCAGCGCAATTAAGCAGCTGGAAACACATCGGCAACAGCTTGATATGCTGTT 1190

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180

Db 1191 TTCTCTCTTGGCGATCAGGTCACTCTGTAGAAATAGCTAAACAGGCTATTGAGAAAGCC 1250

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200

Db 1251 AAGGAAGAACATTATGACTACGTCATTGATGACGCGAGGCGCTTGCATATCGACCAT 1310

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220

Db 1311 GAACGTGATGATGAACATGACCAACGTCGAAGAAATCCGGAATCCGGAAGAAATTTCCCTG 1370

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240

Db 1371 GTGTGCGATTCAATGACCGGTCAGGACGCTGATGATGTTGCCAAAAGCTTTAATGAACG 1430

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAla 260

Db 1431 CTCGGTTTAAACCGGTGTGTGTGTTGACTAGCTGGATGGAGACACACGCGCGGCGCTCG 1490

```
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValIleYMetSerGluLysLeu 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1491 CTTTCTATTGCGCGTGTACAAACAGCGCAATTAAGTTTGCAGGTTTGGCGAAGAGCTT 1550
QY 261 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1551 GATCGGTTAGAGCGGTTCCATCCTCGAAGCGATGCATCAAGGATTCTCGGATGCGGCAC 1610
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnLysAlaLysAspLeu 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1611 GTCGTGACATTGATTGAAAACACACAGCCGCGTGTATGAGACAAACACCAAGAGCTG 1670
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1671 GAACAAAAATGAGACGATGAGCTTCACATGTCGACGATTTCTGGAGAGCTCGGCCAA 1730
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1731 CTCAGAAACATGGGGCGCGTGTATGAGCTTCTGCAAAATGATGCGGGTGCAGGTAAATG 1790
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1791 AAGGCCCTGAAAAACATGCAAGTTGATGAAAAACAGCTGAATCATGTGGAGCAATCATC 1850
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1851 AAATCAATGACTGTTCTTGAANAAGACAGCGCGATATTATCATGCCAGCGCGGGAAG 1910
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1911 CGGATTGCAAAAGAAAGAGCGGACATCCGTGACAGGAAGTCAACCGTCTGCTTAAGCAGTTT 1970
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1971 GATGAATGAAAAAATGATGATGAGCAGATGACAAACATGTCAAACAGGCAAGAAA----- 2024
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2025 -----AAAGGTTTAACTACTCTTTT 2045

RESULT 11
BSUB0009
LOCUS
DEFINITION
  Bacillus subtilis complete genome (section 9 of 21): from 1598421
  to 1807200
ACCESSION
  Z99112 AL009126
VERSION
  Z99112.1 GI:2633902
KEYWORDS
  .
SOURCE
  Bacillus subtilis.
  Bacillus subtilis
  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
  1 (bases 1 to 208780)
  Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G.,
  Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borcherdt, S.,
  Borriss, R., Boursier, L., Brans, A., Braun, M., Brignelli, S. C.,
  Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V.,
  Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J.,
  Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoit, A., Ehrlich, S. D.,
  Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E.,
  Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
  Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Gollightly, E. J.,
  Grandi, G., Guisepi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R.,
  Heaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F.,
  Itaya, M., Jones, L., Joris, B., Karamata, D., Kashihara, Y.,
  Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
  Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
  Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H.,
  Masuda, S., Mael, C., Medigue, C., Medina, N., Mellado, R. P.,
  Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
  Ogawa, K., Ogihara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M.,
  Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P.,
  Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
  Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,
```

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Serron, S. J., Serron, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, P., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, I., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Danchin, A.

The complete genome sequence of the gram-positive bacterium

Bacillus subtilis

Nature 390 (6657), 249-256 (1997)

98044033

9384377

2 (bases 1 to 208780)

Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

Direct Submission

Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr

Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45

68 89 48

Location/Qualifiers

1. .208780

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

258. .4588

/gene="bpr"

258. .4559

/gene="bpr"

/EC_number="3.4.21.-"

/note="alternate gene name: bpf"

/codon_start=1

/transl_table=11

/product="bacillopeptidase F"

/protein_id="CAB13404.1"

/db_xref="GI:2633903"

/db_xref="SWISS-PROT:P16397"

/translation="MRKTKNRLISSVLSTVYVISSLLPFGAGASSKVTSPSYKKELQ
SAESIONKISSLSKKSFKKKTKFLIKFRDLANPEKAAKAAVKKSKLSAAKTEY
QKSAVSVSLYKTDADSQDQVLYKTNQKGNADQIHSYTVVNGIAVHASKKEVMEKY
TQVPEVKVLPNEKROLQPKSSPYNKKAKAKATKATGVWNVQDIDAPKAWALGVDG
TGVFVAS IDTVGNHNPALKKEYKGYNPENPENENMNVDAVAGEPYDLAGT
HYVTGVSSEPDGTNOICVAPGAKTAVKAFSEDDGTADILEAGVWLPKADENCP
HPEMADPVNNSGGSGLEWTFDMWNAARADIFFEPSRGNFDLFIPGPGSLANP
ANYPESFATGATDINKLADFSLQGPSPYDEIKPEISAPGNIRKSVPGTIEDWDG
TSMAGPHVSAVAALKQANASLVDEMEDILSTAEPLTDTFFDPSNNGYHGLVNA
FDVAVTDLGLKAGQVSEGDDEPPYQHEKVTAEYEGSLPLTITAEADVST
VLSYKLDQGEWTEITAKRISGDLKGTQAEIPDKTKLSYKWMHDFGGHVSDD
VYDVTVKPSITAGYKQDFETAPGGWASGTNNWGVSTPNTAASGKVVGTNLT
GNTANSANNLVPPIKAPDGSGLFQPKSNHLEDDPYGVFVLPPEGKWEQAGV
INKTGSWDEBIDLSATKGNQIQVNFQNSDESIAKGGWIDDVVLSKASAKTVYK
NKGVLGSGKGGKKEKPNPKKAPSAVTKHONKAIQVLPKRAQVSVVETGKSTY
SDQSTQYTLKHKAGDTLMAEAYGQSKTKYSLTKDTQTQAMFTLEMKKTKLGT
VINKTGPEVTCASVYVVEDAAVPEAMTNDKGYMELEYEGAYTIVAAPYTSDEF
VELKGDVTVETALKPFVGYPCGEIAYDGTAEANASVTEAGNGWAKVMTLADGKGM
TGLGFWDVETEPFGGTEFFVYDADKCGPKKTAGPNAELNGWVMDLS
SKGMWQDFILYIQSPDFLSPGLAMDGQNGSRNQWIDKQWPGDKACGNMTI
RALVDYAAVPEPTSPDKSTNKSIVYKGNASPGTVHINKEKAGETKAAABGT
FHAGIILNGENELTATSDNGTIDASSITVTLDOEKPLTLDNPKDGGTKRKL
TVKGVSDNLKADKVKNGKATVADGSYSARILLNGRNEIKVATLACNKTTKTV
IDVNFDPKVISLIPGEDKNAKAGESVKIAFSSAEDLATFTIRPLTNRAVSQVAT
ELFLRISPGRYEGVWTATSSIKAKGKVEIVRDDYGNTEKTNKGLNMNTEN"

terminator

4571. .4588

/gene="bpr"

4754. .5683

/gene="spoIIIGA"

4754. .5683

/gene="spoIIIGA"

/function="processing of pro-sigma-E (SpoIIIGB) to active

sigma-E (stage II sporulation)"

/codon_start=1

```
/transl_table=11
/product="protease"
/protein_id="CAB13405.1"
/db_xref="GI:2633904"
/db_xref="SWISS-PROT:P13801"
/translation="MKIYDVIWLNFCEFDALLLLTATYLKRHYKRRVLGVGAFIGS
SIVLLMPTSPIVEHPAGKLAFSVIVVYVTFGFKRFFQNLFSFYFATFLMGGI
IGARSLQSNVQNGVMITQTFGDPISWLVFVGGFPALFFSKRIEDIEYKNIQ
YEYVSQAOLDSOLHVRGLDSGNOLYDLTKTPVMIYIDKLELPFGTATMLIR
NTDPLEALEOLDSPFELDKMLLPIYSGVGQOQFLLCVKPDPHVTIATKEMISADKC
LIGISTKLSADGFDALIHKMLSGRAVRHS"
gene
5746..6465
/genes="sigE"
CDS
5746..6465
/genes="sigE"
/function="early mother cell-specific gene expression"
/notes="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKKLRLTHLWKLMLKGLKSDVEYVIGSEALPPISKDEE
QVLLKLPNGDQAARAILIERLLRVYIARKPENTGINIEDLISIGTIGLIRAVNIF
NPEKKIATYASRCIENEILMYLRNNKIRSEVSEYFENIDMDGNELLISDVIGTD
DIIITKDIEANVDKLLKALEQLNREKQIMELRFLGVGEEETQKDQVDMVGISQS
YISLEKRIIKRLRKEFNW"
gene
6605..7387
/genes="sigG"
CDS
6605..7387
/genes="sigG"
/function="late forespore-specific gene expression"
/notes="alternate gene name: spoIIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSRNKVEICGVDSKLPVLKNEEMKRLFRQLQDEGDSAREKIV
NGNRLVLSIQFNRRGEYVDLDFQVCGTGLMKSDINFDLSHNKFSYAVPMIGE
IRRYLDNNEIRYRSLEDIAYKALQVERLISSETSKEPTAEDIAKLVLPHEIYFA
LDADQPSVLFPIYNDGDPYVADQISDERNTDSOWTEELAKKEGMRRLNDRKMI
LRKFFGCKTOMEVAEEIGISQAQVSLEKAAIKQKNKNIHQ"
gene
7335..8496
/genes="ylnA"
CDS
7335..8496
/genes="ylnA"
/function="unknown"
/notes="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPTREMBL:O31723"
/translation="MILQDNLVSLKRNKGWLKD IHKVEEKENWVLYGLNGAKTAL
LNLCSYFPTSGEMVYLGHEFGTELGKLRKIGLSAALQOKLYPADSAPFIALS
GAYASIGLYETPSKTEREKAIGLLEDGAEYADRIEYLSOGKORALARAALMDP
ELLILDSPVGLDFTAREKLLDITIYIANKENAPSLIYVTHAEELIEVFDKALLKQ
GEVPSGEIKEMITDQILSAFEDTPPIHVLNDRPPLTRAEPITNA"
terminator
8472..8496
gene
8531..9811
/genes="ylnB"
CDS
8531..9811
/genes="ylnB"
/function="unknown"
/notes="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11
```

```
/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPTREMBL:O31724"
/translation="MDQQLISLRKVEEKEELIQLAKTLISYQTPAPPARNTEGIGS
WIAGYLNELGFSIDKWDVYPGDPNVVGLKGTDSADYYSLIINGHVDVAEYKDEEWK
HDFPHPTERKGLLIGRGSMDKMGACVLAFLIRAEATLPGDLILQSVIGVEEWE
AGTLECKRKYHADPAIVADTSDMHIOQGGVITGIEWIKSSOTPHGTRRNMIHAGG
GTFGASAIKEMAKIIAGLGELEHWSIMKSPGFKPCTNTINPAVIEGSRHAFAIDE
CRLLTVHPYNETHDOVAEIEDYVNLSDSDIHLRENRPFVKGSGSSMIEDRGELF
PALEVDGHPGVLAITASHQVKRECPILIDVSQSDVGGWGLDAGIPCVIYVGGDLHN
AHSNKEKVSIEQLVETKLIILDFIISWCSRKEQ"
gene
9894..10139
/genes="ylnC"
CDS
9894..10139
/genes="ylnC"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPTREMBL:O31725"
/translation="MISISEFQVKDVVNSGKLGISGIDIDINVTGKIQAILLGN"
Alignment Scores:
Pred. No.: 2.38e-89 Length: 208780
Score: 1647.00 Matches: 319
Percent Similarity: 84.62% Conservative: 66
Best Local Similarity: 70.11% Mismatches: 60
Query Match: 72.43% Indels: 10
DB: 1 Gaps: 1
US-09-943-108A-2 (1-455) x BSUB0009 (1-208780)
Qy 1 MetAlaphecluglyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 73153 ATGGCATTTGAAGATTAGCCGACCGACGATTCCTAAATTCGCGGAAAA 73212
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 73213 GGGAAAGTCAGCGACAGATGTAAGAGATGATGCGTGAGGTCCGCTTGGCTGCTT 73272
Qy 41 GluAlaAspValAspPheLysValValLysGluPheLleLysThrValSerGluArgAla 60
Db 73273 GAGGCTGACGTTAACTTTAAAGTAGTCAAGGATTTGTCAAAAAGTAAGTGAACGCGT 73332
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLleLysLleValGln 80
Db 73333 GTAGGCCAAGACGTCATAAAGTCTGACGCCGCCGACGAGGTCAATAAAGTTGTCAA 73392
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 73393 GAGAACTGACTGAGCTGATGGCGGCGGAGAGACGACAAATCGCGCTCGCAAAAGCGCG 73452
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrThrAlaGlyLys 120
Db 73453 CGCACTGTTATTATGATGCGGTCTCCAAAGTCCGCGGTAAACGACACACGCGTAAG 73512
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAspIle 140
Db 73513 CTTGCCAATCTCTGCGCAAAAGCATTAATCGCAACCGATGCTGGTGTGCTGCCGATAT 73572
Qy 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 73573 TACCGCCGACGCGCAATTAAGCAGCTGGAAACACACCTCGCAACAGCTGATATGCTGT 73632
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 73633 TTCTCTCTTGGCGATCAGTCAAGTCTCTAGAAATAGCTAAACAGGCTATTGAGAAAGCC 73692
Qy 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisLleAspGlu 200
Db 73693 AAGGAAGAACATTATGACTACGTCACTTTTGGATACGCGAGGGCGCTTGCATATCGACCAT 73752
```


FEATURES		Location/Qualifiers	
source	1.	303249	
		/organism="Bacillus halodurans"	
gene		/db_xref="taxon:8665"	
		/note="alkaliphile"	
CDS		104..1102	
		/gene="BH2267"	
		104..1102	
		/gene="BH2267"	
		/codon_start=1	
		/transl_table=11	
		/product="penicillin-binding protein"	
		/protein_id="BAB05986.1"	
		/db_xref="GI:10174887"	
		/translation="KKNHLHIMEDWKLSTGLMKXGEDIPIASLGFANPARBIPNE HHTFGASCKLFTAIACIQVRAKLSFDLPISDLWLPAPFPNTIHLITHTSGVP DYDEETIDFEDLWKMVYHLRLKDFLPFOHAPMKFPFGPHFYNNAGTILGL VYESVTFQEVYEVANFYQAGHESGYFAFTLPKATLGYIDLEGWKNLYSL PVIGSGGAVTAEDMKLWALMRHLENNETITOKLLTPHVCDDDDYGVGWIK QDGAISKVMGYDPGVCFSAFPTSGNIVVVCANQSSGAYDVMAAIEALISEA"	
gene		complement(1146..1469)	
		/gene="BH2268"	
CDS		complement(1146..1469)	
		/gene="BH2268"	
		/note="BH2268"	
		unknown"	
		/codon_start=1	
		/transl_table=11	
		/protein_id="BAB05987.1"	
		/db_xref="GI:10174888"	
		/translation="MKVITLGGSTFKFRQEQASAYLTILKGVLSVGFQSGIEI TEEQOVFENIHRKTEKSDIEIFLVDVGYIGRSTRKEIEFATSKGMPRIYSESLA STFS"	
gene		complement(1636..2373)	
		/gene="BH2269"	
CDS		complement(1636..2373)	
		/gene="BH2269"	
		/note="BH2269"	
		unknown"	
		/codon_start=1	
		/transl_table=11	
		/protein_id="BAE05988.1"	
		/db_xref="GI:10174889"	
		/translation="MVKVLTVGVFALSTGLMACNGQBEVEKNIDGASGEIPSI BEIIDNNLPDPAVTNGESIDKETVYVLEQQMMWAOYGISLGEKGOMLSLEE QTIFILNELLQOATKEIEVSEEDIELAFITVAQSSSEALMALEKRGSSMDE LREIEHYVQKQYVEETETAINTEEEIOARYEEBKQYITTEDELPFTELQSNIEQR LIAKEQEOLVLEFKLRDSDITVHI"	
gene		complement(2582..3367)	
		/gene="BH2270"	
CDS		complement(2582..3367)	
		/gene="BH2270"	
		/codon_start=1	
		/transl_table=11	
		/product="ABC transporter (ATP-binding protein)"	
		/protein_id="BAB05989.1"	
		/db_xref="GI:10174890"	
		/translation="WADILQVEQNTLRKAKKPILOISNQVSEQEHWALVINGSGKT SLLILITGWPTEGHVAVLGNRYGHPVIGEVKRGVYMSLIDRFRTHKQDTVLEI VLSGLHTEVTEPGEADVAKARALETGLTHLTDRLPFLTSQGERKRAFLARAW AKPELFDTEPTGLDITAROLLQTLLESVATTGAPTEFLYTHYPEIOPFTTHVM LKDGHVTAQGRKDEVLTAKMITAAFGCSLDVTAQSGRYVWKPN"	
gene		complement(3378..4310)	
		/gene="BH2271"	
CDS		complement(3378..4310)	
		/gene="BH2271"	
		/codon_start=1	
		/transl_table=11	
		/product="dipeptidase"	
		/protein_id="BAE05990.1"	
		/db_xref="GI:10174891"	
		/translation="MRIFADHCDALLKLWEDPAREYTDPAEANAVALMEGVSRAQC FAIFVGNIPFDKFOVALQAELEFHEKILKHPQSFIERWDDLLLTLDADQIGAILT	

LEGADPFDDPTKLTLYRLVRSILGTWNOANLCADGVGPRGGGLTEWKAASVLSN
NEHVLTDYSHLSERGFWDVLEYAOPFVATHSNKAICKHRENTDQALKALIEAGGF
IGIVFPHLELTGTDSDASIDDVYKHIHIEICSLGCKHVGFSGDFGICFVDGLEHAGR
FQWMDLLKRYDEELVRGFANDNFKRATPISTV"

4406..4840
/gene="BH2272"

4406..4840
/gene="BH2272"

/codon_start=1
/transl_table=11
/product="heat shock protein class I (low molecular
weight)"

/protein_id="BAB05991.1"

/db_xref="GI:10174892"

/translation="MSREESKSLPNVYQDFLRISDDFFSKSQHLHDMGFAPSI
IRYERKGFILIEALPGVSKNQIQLDVFPQYVIRIHIEEVLREMRDEKGIYSKIS
SQARERIVFPVQEGDKARLNGLLTITAPNHRKRINP"

4864..5193
/gene="BH2273"

4864..5193
/gene="BH2273"

/note="BH2273"

unknown"

/codon_start=1
/transl_table=11
/protein_id="BAB05992.1"

/db_xref="GI:10174893"

/translation="MINLHPQWTWSQAYIQPFQHSHTWPNPQPTSSRGQMYTGTG
PSFPFPFYPFQSKPKASPKAAFTNGDFTDINKTFQTVQVYVTVNQVPLVKS
VGSIFIK"

complement(5202..5510)

/gene="BH2274"

complement(5202..5510)

/gene="BH2274"

/note="BH2274"

unknown"

Alignment Scores:

Pred. No.: 2 43e-85 Length: 303249

Score: 1584.00 Matches: 307

Percent Similarity: 83.33% Conservative: 73

Best Local Similarity: 67.32% Mismatches: 70

Query Match: 69.66% Indels: 6

DB: 1 Gaps: 2

US-09-943-108a-2 (1-455) x AP001515 (1-303249)

Qy 1 MetAlapheGlugLyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||
Dd 210434 ATGGCGTTTGAAGAGTTAGCGGCGCTGCAAGATACGTTGAAGAAATTCGCGGAAAA 210375
Qy 21 GlyLysLeuThrGluAlaAspLleLysLleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||
Dd 210374 GGAAGAGTAAGTGACACAGACGCTGAAGAAATGATGCGTGAAGTTCGTGCGCTCCCTA 210315
Qy 41 GluAlaAspValAsnAspLysValLysGluPheLleLysThrValSerGluArgAla 60
|||||
Dd 210314 GAAGCGGATGAACCTTAAAGTGTGAAGCAATTCATCGCCATGTCGAAGGAAAAAGCC 210255
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLleLysLleValGln 80
|||||
Dd 210254 CTCGGTCAAGAAGTCATGAAAAGCTGACACCTGGACACAGGTTCATCAAGTCGTCAT 210195
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLleAsnMetSerAsnLysPro 100
|||||
Dd 210194 GAAGAGCTGACGGCGCTGATGGAGGCGACAAAGAAATTCGGTTCGAGCAAGCGG 210135
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||
Dd 210134 CCACCGGTCGTGATGATGTCGGTCTTCAAGGGCTGGGAAAAACGACGACGCGGCGAAA 210075
Qy 121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspLle 140
|||||

Db 210074 CTTGCCAACCATTTGGGAAAAACACAAATCGGAACCCGATTCTCGTCGTCGCCGATATT 210015
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLeuGlnIleAspIleProVal 160
 Db 210014 TATCGTCGCGCTGGCATTAAACACCTTGACACATTAGGTAAACAATTAACATCGCTGTG 209955
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 209954 TTTAGCCTGGTGATCAAGTAGCCCTGTGTAATTCGCAACAGCGCATCGACGGCG 209895
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200
 Db 209894 AAAGACAGCATCACAGCATATGTGCTCATGCATACAGCTGGAGCTTCATCATCGATGAA 209835
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 209834 GAGCTGATGATGAATCGACAAGTAAGAGAAAGTGGCAAGCCAGATGAAATCTTGCTC 209775
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240
 Db 209774 GTCTTGATCGGATGACAGGGCAGATGCGGTTAATGTGCGGAAAGCTTCATGAGCAG 209715
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 209714 CTCGATCTTTCGGCGCTGCTTTTGACGAAGCTGGATGGGATACTCGAGGCGGAGCTGCC 209655
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
 Db 209654 CTCFCGTTAAGCGGTTACGAACACGCCGATCAAAATTTGCCGGATGGGTGAAGAAGATT 209595
 QY 281 AspGlyLeuLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 209594 GACCAGCTGAACCTTCACCCAGAGCGAATGGCTTCGGGATCTTGGGATGGGGAT 209535
 QY 301 ValLeuSerLeuLeuGlnLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
 Db 209534 GTGCTGACATCTCAATGAAAAACCAATCCACTGCGATGAAGAAAAAGCGCGTGAACCTT 209475
 QY 321 GlnLysLysMetArgGlnSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
 Db 209474 GAGAAAAAATCGGTACAATGACATTACGTTTCGATGATTCTTGAGCAACTCGACCAA 209415
 QY 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 209414 GTGCGAGCATGGACCATCTCATGACTGCTCGAATGATGCCAGGATGAACAAGCA 209355
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAlaIle 380
 Db 209354 AAAGGCTTGAATACTTGAAGTGCAGCAGAGAACAATTAAGCCGGCTTGAGCGCATGTGC 209295
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 209294 GTTCCATGACGAAGCAAGAGAACAATCCTAGCATCAATTAATGCGAGTCGTCGCGA 209235
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
 Db 209234 CGGATCGGAAAGAGAGCGGGCGCGGATTCAAGATGTCACCGCTGTTAAGCAGTTT 209175
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGly---GlyGlyLysGlyLysGly 439
 Db 209174 GAAGAGATGAAAGATGATGAAGCAATGACGAACATGACGAGCAAGCAAGCAAGAGAA 209115
 QY 440 LysArgAsnGlnMetGlnAsnMetLeuLysLysGlyMetAsnLeuProPhe 455
 Db 209114 GGGAAAGGC-----TTTGGAGGCTTTAAATGCGGTTT 209082

RESULT 13

AX433432

LOCUS AX433432 1136 bp DNA linear PAT 28-JUN-2002

DEFINITION Sequence 1847 from Patent WO0291113.

ACCESSION AX433432

VERSION AX433432.1 GI:21658236

KEYWORDS

SOURCE Bacillus licheniformis.
 ORGANISM Bacillus licheniformis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 AUTHORS Berka,R. and Clausen,I.G.
 TITLE Methods for monitoring multiple gene expression
 JOURNAL Patent: WO 0229113-A 1847 11-APR-2002;
 NOVOTRIZ Biotech, Inc. (US) ; NOVOTRIZ A/S (DK)
 FEATURES
 source location/Qualifiers
 I.1136
 /organism="Bacillus licheniformis"
 /db_xref="taxon:11402"
 BASE COUNT 338 a 263 c 326 g 209 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,68e-77 Length: 1136
 Score: 1402.00 Matches: 270
 Percent Similarity: 87.30% Conservative: 60
 Best Local Similarity: 71.43% Mismatches: 48
 Query Match: 61.65% Indels: 0
 DB: 6 Gaps: 0
 US-09-943-108A-2 (1-455) x AX433432 (1-1136)
 QY 1 MetAlaPheGluGlyLeuSerGluArgGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGCATTCGAAGGATTTACCCGCGACTGCAGCAACAGCATCTCAAAATCCGCGGAAA 60
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGAAAGTAAACGAGCAAGCATCAAGAGATGATGCGGAAAGTCGTTAGCGCTCCTT 120
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
 Db 121 GAAGCGAGCTCAATTTAAAGTCGTGAAGATTTGTGAAAGGTAAAGCAACGAGCT 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 Db 181 GTCGGACAGACGCTTATGAAGCCCTGACGCCCGCAGCAGCATTAATAAGTGGTAAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAAGAGCTGACGAGCTGATGGCGCGCAGGAAAGCAAGATTGCCGCTGCAAGAGCTCG 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CGACCGCTCATCATGATGTCGCTCTGCAAGGGCGCGTAAACAGCAGCTCATATGCGGAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 CTGCGCAACCTGTTGCGAAAAACATTAACCGCATCCGCTGCTGCGCAGCAGCATC 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TACCGCGCGGCGCATCAACAGCTGGAACACGCTGCGCAAGCAGCTCATATGCGGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TTTTCACTGGGTGACAGGTAGTCGCTGCGTAAAGCGCGTAAACAGCAGCCCATCAAAAGCG 540
 QY 181 LysGlnGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200
 Db 541 AAAGAGATCATCACGATACCTTCTCATCATGACGCGGCGGACGCTTCATATTCATGAA 600
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 GAGCTGATGATGAGCTTGAGCAGGTAAAGAACCCCTCAGCCGGAAGAGATTTCCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240
 Db 661 GTCGTTGACTCCATGACGGGTCAAGACGCGCTCAATGTTGCCAAAAGCTTTAAACGACAG 720

Qy 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAla 260
Db 721 CTGGTGTGACAGTGTCTGACGAGAGCTTGACGCGTACTAGAGCGGAGCGCC 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
Db 781 TTGTCAATCGCGCGGTGACAAACACGCGGATCAAAATCGCGGAAATGGCGAAAAGCTT 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGCGCTGAGCGCTTTCATCTGACGAGATGGCTTCGCGGATACGCGAATGGCGCAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGCTACCCCTATCGAGAAAGCCAGCCACGCTAGACGAGAAAGGCGAAGGACTC 960
Qy 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
Db 961 GAGCAAAAATGAAAAATATGAGCTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCCGCAATATGGCCGCTTGAGACCTGATTCAAATGATGCCCGCGGAGCAAAATG 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378
Db 1081 AAAGGCTTAAAAAGCTGAAGTTGAGTAAACACAGCTCACCCATATCGAAGCG 1134
RESULT 14
AE007684 11024 bp DNA linear BCT 27-JUL-2001
LOCUS Clostridium acetobutylicum ATCC824 section 172 of 356 of the
DEFINITION complete genome.
ACCESSION AE007684 AE001437
VERSION AE007684.1 GI:15024715
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 11024)
AUTHORS Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,K., Jee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
Tatsov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
TITLE Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
11466286
REFERENCE 2 (bases 1 to 11024)
AUTHORS Childress,D., Zeng,Q. and Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
source
location/Qualifiers
1. 11024
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:11488"
84. 317
/gene="CAC1747"
/note="acpA"
84. 317
/gene="CAC1747"
/codon_start=1
/transl_table=11
/product="Acyl carrier protein, ACP"
/protein_id="AAK79713.1"
/db_xref="GI:15024716"
/translation="WVFKEVKDITADQLGIDATEIKNESSFIDDLGADSLDIYELIMA

LEEFEDIEMPEAEAKVSSVGVVYVIKAHTEE"
444. 1136
/gene="CAC1748"
/note="rncS"
444. 1136
/gene="CAC1748"
/codon_start=1
/transl_table=11
/product="rRNA-specific ribonuclease III"
/protein_id="AAK79714.1"
/db_xref="GI:15024717"
/translation="MEERLEELLESKLGTEFONINLLVTALTHSSVANENKNAEYNE
RLFLGDAVLQLSISEVFFKYPPTISGELTKKRALVTCGMSLHISGERQLCKYIRM
SHEELTGGTRWYSIIADCEAVAAIYLDKGFDAKAFILREPEGICIQNAVENKILL
DAIKRLQELIQSKGDKIKITLVRHGEPFHRKFFVNLNFDNDVKSTGEGITTKDAEQ
DAKALKKGLDN"
1137. 2213
/gene="CAC1749"
1137. 2213
/gene="CAC1749"
/codon_start=1
/transl_table=11
/product="BioB/LipA-like protein"
/protein_id="AAK79715.1"
/db_xref="GI:15024718"
/translation="MGKSHYIPIFVPEHCGPDCVFCNONTITGEESKYNVDYVKST
VEDYLDIIPKSAATIEISFGGTTAINIKOKELLSVABYKIGIKYIRLSTPDD
YIDEILTNLKNYSVDIEELGVQLSEVLLASGRGHKEDYVKAASLIIKKYFTLGH
QIMGLPGDGAEDKDIETTRVICMKPNICRIYPALVIKDTAWERKMYNRGIIKPYDTEE
CVTSIKVFMGFANDIEVIRIGLPTEINTDELVAGPFPAPRILEVSGSINMEI
FNDLKESNTVDIYNPKDISKLYSNKKTFFYSMKDKLNLNLRINVIQNPVEVEKDTLK
MFDNCKIVSKYEVLVKYKEG"
2270. 2530
/gene="CAC1750"
2270. 2530
/gene="CAC1750"
/codon_start=1
/transl_table=11
/product="SpoVS-related protein"
/protein_id="AAK79716.1"
/db_xref="GI:15024719"
/translation="MEVLKVSQSPKSNAGALAAVLNRSNSAEVQAVGAGAINQAVK
AIAITGFVAPNGIDLVVVPAPAEISGEDRTAIFKIVEF"
2747. 6322
/gene="CAC1751"
/note="smc"
2747. 6322
/gene="CAC1751"
/codon_start=1
/transl_table=11
/product="Chromosome segregation SMC protein, AtPase"
/protein_id="AAK79717.1"
/db_xref="GI:15024720"
/translation="MFLASIRGRKSFADKTDLLFKNGITAVVPGNSGKSNISDAV
LWLGQSVKNLGRGMEDVIFAGTYRKSQVLAQVSLIDNSDKQNLNLYSEVTVSR
KLVSQSYIYNNTCKRLKDIDELFMDTIGIGESYIIGQCKIEAVLSQKPERRAL
LEAAGIVKFKRYDAENKLENTQNLVINDILFTEERLEPRLSEKARFVTEL
SDELAKEINTIYISDINDYINDLKQRMADKLSIDENVKDEKTSLSKGVATSEL
DEFDKVSNNKTYVESKSEHOKLSIELLEKTSNSDVAKNLYKEIOLDNSIVN
LKSEYELTKTDEKYNKELLSSKINSEKKKIDGLIPEWKSUKOYKNDADII
STSQNNNEVILKEISNESKESIKRAGESYKSLKINEVTKNLSIELYKINDK
ISGYNQIRENSKISKLNRIISDEKLNELNSKNLEANKMLINLEKQYEGYNR
SVKNMQHVTGFDVGPSESSVILGEYVKVKEFEFAYEISLGAISDIITLDDNAK
KLINLYKSNLRATFLPLNIQKRLINISDARHEKGFIGTASELIDYDSTPLAVN
VYLGRTVYDNDMSALKIATKLSYSPKIVTLGTEVNVPGSLTGTSTYKAAISIGRK
RETEELNLNVSQALQSNNKILLENKKVVELNCLDLDTDHEKTELKIKER
LKSIDIESEKLNKSNYSVSGIFGFKIKHLEKLVKEENKALKLRANNNDLID
ELERKLDENSKVNLNEEIMSKVQKAKSDENLMSSTRETERYKVMHNMENKISL
KNEIDDFQNKILANKTYNNGVQIKOIVLELNSFKDSKSEVERKAGNTENRN
SLOGINVLQKLETKHRYEINLAKITESENLYRLNSDFKLITYSAAAEKFEIEM
ISVKKKIDELKRELKMGVNVASIEYKVEKYEKTFPANSQKELDNAKDELIRYLEE
WTSKMTVNEPNFKLNENKILFRDLFGKGNADLILSGDDELASSIEINVEPPGKL
QNINLMSGGEGKLSAIALLLFAILKQKPTFFCILDETEALDDANVARYAEFLRRFSEN

gene 6352. .7263 /TOFIVTHRGKTMASDVLYGVTMEKGYKSVISLDLDSORENVYS"
/gene="CAC1752"
/note="fts1"
CDS 6352. .7263 /product="signal recognition particle GTPase, FtsY"
/gene="CAC1752"
/codon_start=1 /db_xref="GI:15024721"
/transl_table=1 /protein_id="AAK79718.1"
/product="signal recognition particle GTPase, FtsY"
/db_xref="GI:15024721"
/translation="MLGNFFDKLNGSLTKNNLTSTKISDWLSTVITIDDLYLEEE
ILITSDIGVETSLIYIDKLKEIKENKVKDPSLVNDCLKEVIKIDLGDKGSGMPKQI
PTILVGVNGVGTTSIGKISSRLKSNLKVIMAAADTPRAAIDOLEWASKRANVD
LIKHEGSDPAVVDPAIEAAKAKADVLICDTAGRLNNKLMNLEKINLIIDREF
SOSTRTLLVDATGQNAVIOAKQFNEVKVDGVLVILKIDGTAKGQVILFKHOLDI
PVEFVGVEGIDDIQEFNPSDFEALF"
7341. .7691 /gene="CAC1753"
7341. .7691 /gene="CAC1753"
/codon_start=1 /transl_table=1
/product="Uncharacterized conserved protein, YLXM
B.subtilis ortholog"
/protein_id="AAK79719.1"
/db_xref="GI:15024722"
/translation="MEERIYLSMLSTYGSILTEKQVNMVLYDDDLSPETAELENN
TFQAHLIDLRKCHMLNLYENKLVIRYKRNEDIKDKIDKIDLYILKDIQKQNVQ
LIDEIEKDISFSI"
7710. .9059 /gene="CAC1754"
7710. .9059 /note="ifh"
7710. .9059 /gene="CAC1754"
/codon_start=1 /transl_table=1
/product="signal recognition particle GTPase Ffh"
/protein_id="AAK79720.1"
/db_xref="GI:15024723"
/translation="MAFEGSTKLOAMKKLRGKLSKDIKDAMEVKLALLRADV
NVIKNFVYVGVKCLGNEVMESLPQQQVIKIVNEELTNLMGKEESKEFENGIT
VIMVGLQAGTTCMGKLSLSLKKKXPLVACDIYRPAAIKOLEVWCKSDYVPFL
SMGKVPNDVTSKAMKAKENGLNVIIIDTAGRLHIDDQLMNELENIKSEFNPKEIL
LVDSMTGDAVNVASFNDKLBELTGVILKIDGTGCGAALSIREMTGKPIYVGLG
ERNQLEIHPDRMASRLGNGDVLILIEKQASIDERQAKELGDRMLSGEFDLFL
QAFQMKLIGPTGLLEKLVFNSMLKGVDSKNEGEMKYEAIKSMTAKERKNPS
LITSTASRRRIALGSGTTVQEVNKLKLNFEQMKMKQFKGNKFSKGLFGKMPF"
9097. .9342 /gene="CAC1755"
/note="tpsp"
9097. .9342 /gene="CAC1755"
/codon_start=1 /transl_table=1
/product="Ribosomal protein S16"
/protein_id="AAK79721.1"
/db_xref="GI:15024724"
/translation="MKVIRLRMGAKKAPFYRIVVADSRSPRDKGFIEELGYNPTE
PVTKVDADKVNAMKNGAOPSETVKKLLDKSGVTTK"
9363. .9590 /gene="CAC1756"
9363. .9590 /gene="CAC1756"
/codon_start=1 /transl_table=1
/product="Predicted nucleic acid binding protein,
containing KH domain"
/protein_id="AAK79722.1"
/db_xref="GI:15024725"
/translation="MKQLEETIAKSLVDCPDEVQYSEVTGQSGSIILEKVPEDMGK
IGKQRIAKAIRTVIKAAAVKENKRVVVEII"
9718. .10218

CDS /gene="CAC1757"
9718. .10218 /gene="CAC1757"
/codon_start=1 /transl_table=1
/product="RimM protein, required for 16S rRNA processing"
/protein_id="AAK79723.1"
/db_xref="GI:15024726"
/translation="MEDEFSIGQILNHGVGELKIYPLTDDINRFDDLDSSVVDNEI
KVISVKKQPNKLIKLEGIDTLDRAVKYKIKVREDAVELKQGIKIDIKGICN
VEDENKQDLGEYDVIYSTKNDVYIKRKGQDILVPALKDILVKIEINKKIVIRAV

Alignment Scores:
Pred. No.: 2..2e-75 Length: 11024
Score: 1393.00 Matches: 280
Percent Similarity: 76.47% Conservative: 71
Best Local Similarity: 61.00% Mismatches: 94
Query Match: 61.26% Indels: 14
DB: 1 Gaps: 5

US-09-943-108a-2 (1-455) x AE007684 (1-11024)

Qy 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 7710 ATGGCTTTTCGAGGATTATCTACAAAATTACAGCAGCAATGAAAAAACTCAGAGGAAAA 7769

Qy 21 GlyLysLeuThrGluAlaAspLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 7770 GGTAAACTTTCTGAAAAGACATAAAGATGCTATGAGAGAAGTTAAGCTTCTCTTTTA 7829

Qy 41 GluAlaAspValAsnPhelysValLysGluPheIleLysThrValSerGluArgAla 60
Db 7830 GAAGCAGATGTTAACTACAAATAGTTAAGACTTTCTTAAAGTAGTAGGAGAAAGTGT 7889

Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 7890 CTAGTAATGAAGTATGGAAGCTTACTCTCGGCAGCAAGTTATATAAATAAGTTAAT 7949

Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 7950 GAAGAGTTAACTAACCTCATCGGTAAAGAGGAGGAAAAATAGATTTTGGGAAAAATGGA 8009

Qy 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 8010 ATTACTGTTAATGAAGCAGTGGACTTCAGCTCGAGGAGAAACCACCTATGCTGGAAG 8069

Qy 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 8070 CTTTCTCTTACTTTAAAGAAAAA---AATAAAAAAGCCACTTCTTGTGCTATGATATA 8136

Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 8127 TATAGCAGCAGCAGCTATAAGCAGTTAGTAGTAGTTGGAAGATCATATAGATGCTCCTGT 8186

Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 8187 TTTTCTATGGGGATAAGTTAATCCTGTTGATATTTCAGAGCTGCATGAGCATGCT 8246

Qy 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 8247 AAAGAAATGGTTTAAATGTAGTTATAATAGATACTGCGGGAAGACTTTCATATTGATGAT 8306

Qy 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220
Db 8307 CAGCTTATGAATGAAGTGAAGACATAAAAGTGAAGTTAATCCTAAGGAAATTTTGCTT 8366

Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 8367 GTTGTGTTATTCATGACAGGTCAAGATGACGTAATGTTGCTGAAAGCTTTGATATATAG 8426

Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 8427 CTTGAAGTACTTGCTGTTGCTTACTAAACTTGATGCTGATACAAAGAGTGGTGTCTGCA 8486

Search completed: February 25, 2003, 02:59:00
Job time : 4087 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

PD 17-MAR-1999.
 XX 24-AUG-1998; 98EP-0306741.
 XX 10-SEP-1997; 97US-0927216.
 XX (SMIK) SMITHKLINE BEECHAM.
 XX Wallis NG;
 XX WPI; 1999-169238/15.
 DR P-PSDB; AAY00910.
 XX
 XX New Staphylococcus aureus Signal Recognition Particle (SRP) with
 PT protein (fih) and RNA (fif) components - the SRP gene and protein
 PT useful as diagnostic reagents and for prevention and treatment of
 PT Staphylococci infections which cause otitis media, septic arthritis
 PT and toxic shock syndrome
 XX
 XX Claim 1: Page 28; 35pp; English.
 XX
 XX This sequence encodes the Staphylococcus aureus signal recognition
 CC particle (SRP) fih component. Fih polynucleotides are useful for
 CC diagnosing a disease related to expression of fih polypeptides by
 CC analysing for the presence/amount of fih protein in a sample due to
 CC infection of a micro-organism with the gene, or determining the nucleic
 CC acid sequence encoding fih. Fih polypeptides and antagonists are useful
 CC for treatment of an individual in need (polypeptide) of, or needing to
 CC inhibit (antagonist) fih polypeptide levels. Fih polypeptides and
 CC polynucleotides are useful for identifying agonists and antagonists by
 CC binding and observing the affect of fih polypeptide activity, which are
 CC potential anti-microbial agents. Fih polypeptides and antigenic fragments
 CC are also useful for inducing an immunological response (T cell/antibody)
 CC to protect against disease, by direct administration (vaccine), or via a
 CC vector (gene therapy). Anti-fih antibodies are useful as antagonists, and
 CC for protecting against disease. Diseases diagnosed, prevented and treated
 CC include those caused by infection, especially bacterial infection,
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound
 CC infection and septic arthritis. Fih polypeptides are antimicrobial, and
 CC are useful for bathing wounds and implants prior to surgical
 CC implantation.
 XX
 SQ Sequence 1368 BP; 521 A; 186 C; 288 G; 373 T; 0 other;

Alignment Scores:
 Pred. No.: 3,47e-197 Length: 1368
 Score: 2274.00 Matches: 455
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 20 Gaps: 0

US-09-943-108A-2 (1-455) x AAX27221 (1-1368)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCAATTGAGCGTTATCAGAACCCCTGCAGGCGACGATCAAAAAAATGCGTGTAA 60
 QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGTAAACTTACTGAGCTGATATTAAGATAATGATGCGGTGAAGTAAGATTAGCGTTATT 120
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheLeuLysThrValSerGluArgAla 60
 Db 121 GAGCGTGCAGTAACCTTAAAGTGGTAAAGAAATTTATTAACAGTATCAGACGCCGA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLleLysLeValGln 80
 Db 181 TTAGGTTCCGATGTAATCAATCATTAACACCAGGCGCAACAAGTTATTAATAAGTTCAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAGGAATTAACGAAGTTGATGGGTGGAGAAATATCATCGATTATATGTCATAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCTACTGGTTTATGATGGTTGGTTTACAGGTGCTGGTAAACAACTGAGGATAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTAGCATTATTGATGCGTAAAAAATACACAAAAAACCCTATTGTTAGTTGACAGATAT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATGATATTCCTGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGATCTAATGATTAATGATTAACATGCT 540
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
 Db 541 AAAGAGACATTTAGACTTTGTAATCATTGATACAGCAGTCGATTACACATCGATGA 600
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 GCATTGATGACGAATTAAGAAAGTAAAGAAATTTGCTAAACCAACGAAATATATGTTA 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
 Db 661 GTTGTCGATCAATGACGGTCAAGATGCTCAATGTTCAGAAATCTTTTGACGATCAA 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 CTGTGATGTCACAGGTGTTTACCTTAACCTAAATTAGATGCTGATACACGTGCTGTCAGCT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGlnLysLeu 280
 Db 781 TTATCTATTCTCTGGTGACAAAAACCAATTAATTTGTTGATGATGAGTAAAGTTA 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 GATGGTTAGAGCTATTCCATCTCCAGCTATGGCATCACGTATTTTAGGTATGGGTGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 GTGTTAAGTTTAATGAAAAAGCGCAACAGATGTGGATCAAGAAAAAGCAAAAGATTTA 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 Db 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTTAGATGATTTTTTAGAACCAACTTGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTGAAAAATCTAGGACCACTGGATGATATATGAAATGATTCAGGTATGATAAATG 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
 Db 1081 AAAGGCTAGATAAGCTTAATATAGCTGAAAAAGCAAAATGATCATATTAAGCGATTATC 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 CAGTCAATGACCGCGGTGAAAGAAACAATCCAGACACATTGGAATGTATCCAGTAAAAAG 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATTGCTAAAGGGTCTGGTCGTTCAATTACAGAGAGTCANTCGTTTGATGAACAATTT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysGlyLys 440
 Db 1261 AAGCATATGAAGAAAAATGATCAACCAATTCACCTGTCGCGTAAAGGTAAGGTAAGGTAAG 1320
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 Db 1321 CGCAATCAATGCAAAATATGTTAAAAAGGTATGAATTTACCGTTT 1365

RESULT 2

AA555331
ID AA55331 standard; DNA; 1368 BP.
XX
AC AA555331;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus DNA for cellular proliferation protein #1643.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PSDB; AAU37472.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 8968; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1368 BP; 517 A; 187 C; 291 G; 373 T; 0 other;

Alignment Scores:
Pred. No.: 5,26e-196 Length: 1368
Score: 2261.00 Matches: 452
Percent Similarity: 99.78% Conservative: 2
Best Local Similarity: 99.34% Mismatches: 1
Query Match: 99.43% Indels: 0
DB: 23 Gaps: 0

US-09-943-108a-2 (1-455) x AA555331 (1-1368)

QY 1 MetAlaPheGluGlyLeuSerGluArgGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATFGCAATTTGAAGGGTTATACAGACGCTTGCAGCAGCAGATGCAAAAAATGGGTGAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACCTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTACGGTTACTT 120
QY 41 GluAlaAspValAsnPhelYsValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCTGACCTAAACTTTAAAGTGTAAAGATTATTTAAACACAGTATCAGACGGCCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTTCGGATGTAATGCAATCATTAACACGCGGCAACACAGTTATTAAATAGTTCA 240
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAGTTGATGGTGGAGAAAAATACGTCGATTAAATGTCAAAATAACCA 300
QY 101 ProThrValValMetMetValGlyLeuGlnIleAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGTTGGTTTACAAAGGTGCTGTTAAAAACAACAACCTCCAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTTGGTGGTAAAGTAAACACAAACAAATTTGTAATGTAATGCAATTAACATGCT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGTCGACGACGATAAATCAATTACAAACAGTAGGAAACAATAATTGATTTCTCTGA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGCAGATCAAGTAAAGCCACACAAATTTGTAATGTAATGCAATTAACATGCT 540
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTGTAATCATTTATACAGCAGGTGATACACATCATGATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGAACGATTAAGAGAGTAAAGACATTCGTAACCAACAACGAATTTATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGATTCATGACGGGTCAAGATGCTGCAATGTTGCAGAACTCTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACCTTAATTAATTAATGATGATACAGCTGGTGGTGCACCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCTTCGGTGCACAAAAACCAATTAATTTGTTGGTATGATGAGTGAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGTTTTAGAGCTATTCCATCTCTGAACGTATGGCATCAGCTATTATTAGGTATGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTTAATGAAAAAGCCCAACAGATGGATCAAGAAAAACAAGAAATTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAAGATGCGTAGTATCGTTACTTTAGATGATTTTTTAAACAACCTTATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCACTGGATGATATATGAAATGATTCACAGTATGATTAATG 1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 ARAGGGCTAGATAAGCTTAATATGAGTGAAGAAAGCAAATGATCATATTAAAGCGATATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACCGCGCTGAAAGAAACAATCCAGACACATTTGAATGTATCACGTAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGCTCGTGGTTCATTAAGAAGTCATCGTTCGTAAGCAATTT 1260
QY 421 AsnAspMetLysGlyMetMetLysGlnPheThrGlyGlyLysGlyLys 440
Db 1261 AACGATATGAAGAAATGATGAACAGTTCACTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProphe 455
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365
RESULT 3
AAS51783
ID AAS51783 standard; DNA; 1365 BP.
XX
AC AAS51783;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus DNA for cellular proliferation protein #200.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PA Haselbeck E, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PSDB; AAU33924.
XX

New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Claim 27; Seq ID No 4365; 51pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC
CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1365 BP; 518 A; 190 C; 285 G; 372 T; 0 other;

Alignment Scores:
Pred. No.: 1,2e-193 Length: 1365
Score: 2235,00 Matches: 447
Percent Similarity: 98,90% Conservative: 3
Best Local Similarity: 98,24% Mismatches: 5
Query Match: 98,28% Indels: 0
DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS51783 (1-1365)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTTGAAGGCTTATCAGAACGCCCTCGAAGCGAGATGCAAAAAAATCGGTGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLysPhe 40
Db 61 GGTAACTTACTGAGCTGATATAAGATAATGTCGTGAAGTAGAATTAGCGTTACTT 120
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCTGAGCTAACTTTAAAGTGGTAAAGAATTTATTAAACAGTATCAGAACGCGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGTTCGATGTAATGCAATCATTAACACCGAGGCAACAAGTTATTAAATAGTTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAGCTGATGGTGAGAAATATATATCGATTATATGTCAAATTAACCA 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGGTGTACAAAGTGCTGGTAAACAAACAACCTGCAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCTATTGATGCGTAAATAATACAAAAACCTTAGTTTGCAGCAGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATTAATCAATTACAAACAGTAGGGAACAATTTGATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAATAATTTGTAATTCATTAACACATGCT 540
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGACACATTTAGACTTTGTATCATTTGATACAGCAGCTCGATTACACATCGATGA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGAATGAATTAAGAAGATTAAGACATTTGCTTAACCAACCAAGAAATTTGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGATTCATCAATGCGGGTCAAGATGCTCAATGTTGCAAGATCTTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
Db 721 CTTGATGTCACAGGTTTACCTTAACFAAATTAGATGGTGATACACAGTGGTGGTGAGCT 780

```

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
DB 781 TTAATCTATTGCTCGTGACAAACCAATTAATTTGGTATGAGTGAACATTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
DB 841 GATGGTTTAGAGCATTCCTCCTGAACGATGCGCATCGATGATTTTCAGGCGCTGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
DB 901 GTGTCAAGTTAATTGAAGAACCGACACAGATGTGATCAAGAAAGCAAAAGATTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
DB 961 GAGAAAGATGCGGAGTCATCATTTACTTTAGATGATTTTGTAGAACAACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspPheMetLysMetIleProGlyMetAsnLysMet 360
DB 1021 GTGAAAAATTTAGGACCACTGGATGATATTATTAAATGATTCAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLe 380
DB 1081 AAAGTCTAGATGAGCTTAATGATGAGTGAAGCAAAATGATCATATTAAGCCATATTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
DB 1141 CAGTCAATGACGCGGCTGAAAGAAACAATCCAGACACATTAATGATATCATCGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
DB 1201 GCTATTGCTAAAGGTTCTGCTGCTTCATTAACAAGAGTCAATCGTTTGTATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
DB 1261 AACGATATGAGAAATGATGAAACATTCCTGCTGCGGTAAAGGTAAATGAAGGTAATAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
DB 1321 CGCAATCAATGCAAAATATGTAAAGGTATGAATTTACCGTTT 1365

RESULT 4
AAS54393
ID AAS54393 standard; DNA; 1368 BP.
XX AC AAS54393;
XX DT 13-FEB-2002 (first entry)
XX DE Staphylococcus aureus DNA for cellular proliferation protein #705.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200170955-A2.
XX XX 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX XX (ELIT-) ELITRA PHARM INC.
XX PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;

```

```

XX WPI; 2001-611495/70.
DR P-PDB; AAU36534.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS Claim 27; Seq ID No 8030; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1368 BP; 520 A; 190 C; 285 G; 373 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1,21e-193 Length: 1368
Score: 2235.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS54393 (1-1368)

```

```

QY 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGGCATTGAAGGCTTATCAGAACGCTTCAGACGACGATGCAAAAAATGCGTGGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspPheLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGTAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTACTT 120
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
DB 121 GAGGCTGACGTAACTTAAGTGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DB 181 TTAGGTTCCGATGTAATGCAATCATTAACACGCGGCAACAAGTTATTAATAATAGTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DB 241 GATGAATTAACGAAGTTGATGGTGGAGAAATATATCGATTATATGCAAAATAAACCA 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CTAATCTGTTGTTATGAGTGTGTTTACAGGTGCTGGTAAACACAACTCCAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrsLysLysProMetLeuValAlaAspIle 140
DB 361 TTAGCATTTATTGATGCGTAAAAAATACAAACAAAACCTATGTTAGTTGCACGACATTT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
DB 421 TATCTCCAGCAGCGATAAATCAATTACAAACAGTAGGAAACAAATTTGATTCTCTGTA 480

```



```
Db 188164 GAAGAATTAAACAGCCTCATGGCGGAGAGAAAGTAAAAATCGGAACAGCTGATCGCCG 188223
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 188224 CCACCGTTTATGATGGTAGGTGTACAGAGCTGGTAAACAACCACTTCAGGAAAA 188283
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysPrometLeuValAlaAlaAspIle 140
Db 188284 CTCGCTAATTATTACGTAAAAATATATATCGTAAACCTTTACTAGTCGCAGCAGATAT 188343
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 188344 TACCGACCTCGACCAATCAACAACTAGAAACACTGGCAACAATTAGATATGCCAGTA 188403
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 188404 TTTTCTTAGGGGATCAAGTAGCCCAAGTAGAAATCGGGAACAGCTATCGCTAAAGCA 188463
Qy 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 188464 AAAGAAGAACATTTAGATTATGTCAATTATCGATACAGCTGGCTCTTCATATCGACGAA 188523
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 188524 ACTCTGATGGAGCAATTAACAACTAGTAAAGAAATCGCTACGCCCACTGAAATTTACTT 188583
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 188584 GTAGTTGATCAATGACTGGCAGAGCGCAGTAATGTGGCCCAAGCTTCACGAAACA 188643
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
Db 188644 TTAGAATTAACGGCGCTGTATTAAACAAATAGACGGTGATACACGTGTGGGGCAGCA 188703
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 188704 CTTTCCATCCGTTCACTAGTCAGCAAGAAACCAATCAATTCGTTGCTACCGGTGAAAAATG 188763
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 188764 GAAGCAATCGAAACCTTCATCCGATCGTATGGCTTCAGAAATTCGGCATGGGTGAT 188823
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 188824 GTACTTTCCTTATTGAAAAAGCACAACTGATGTAGATGCAGAAAAAAGAAAGCTATG 188883
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 188884 GAACAAAAAATGAAAGACACAGCATGACCTTAGATGACTTCTTGGACCAATTCGAACAA 188943
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 188944 GTAAACAATGGGACCACTAGATGAACACTATAAAATGATGCCAGGGGCAACACAAATG 189003
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 189004 AAAGGCTCGACACATGCAACTAGATGATAAACAACCTCGGTACATCGTGAAGCGATAAT 189063
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 189064 AATTCATGACCAAAAACAGAAAGATTAATCCGGACATCAATCGCAGCAGCAAGAAAA 189123
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 189124 CGAATTGCTCGTGGAGCGGACGCCCAATTCAGAAATCAATCGCTCCCTTAAACAAATTT 189183
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
Db 189184 GCTGAAATCAAAAAATGATGACCAAAATGACTGGTGGGGAAGGTAAGAAGGTAA 189243
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 189244 -----AATCCATTCGGCAATTTCAAAATGCCATTT 189273
```

```
RESULT 7
ABN92495
ID ABN92495 standard; DNA: 1014 BP.
XX AC ABN92495;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy; gene; ds.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2002-381255/41.
XX DR P-PSDB; ABP39950.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PS polypeptide, useful for diagnosing and treating bacterial infections -
XX PS Disclosure; SEQ ID 1958; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
SQ Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;
Alignment Scores:
Pred. No.: 3.75e-130 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 24 Gaps: 0
US-09-943-108a-2 (1-455) x ABN92495 (1-1014)
Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGCGATTTGAAGGATATTCGATCGCTTACAGCACGATGCAAAAATTCGGTATAA 75
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAGAGTAACAGACAGAGATATTAAAACGATGATGCGTGAAGTGACATAGCGTTATG 135
Qy 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 136 GAAGCGGATGTTAACTTCAAGATGTTTGAAGATTTGTTAAGAAATGTTTCAGAACGAGG 195
```


QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 241 GAAGAATTAAACGAAACGTTAGTTCAGAAACGGTGAACATAAATCTCCAAATC 300
QY 101 ProThrValValMetSerValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 301 CCACAGTGTATGATGACAGGGTTACAGGGCTGTAACAACTTTTACTGGTAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 361 TTAGCAAAACACTTAATGAAACTGAAGACGCTCGCGCTTTAATCGCTGGTGCCT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 421 TATCGTCCAGCAGCATGATCAGTTGAAGTTTGTAGTCAACAATAGAGTCCCGTT 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 481 TTTGATATGGGACACAGATGCTTAATCCAGTGGAAATGTCTCAAGGTTAGCATTAGCA 540
QY 181 LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 541 AAGAAAGAAAGAAATGATTAGTCTTAATGATACGGCGCGCTTTACACATTGACGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 601 GCCTTAATGACGAATTAAGAAATTAAGAGTTGGCTAATCCCAATGAATTCGTGA 660
QY 221 ValValAspSerXerThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 661 GTTGTGTGATCCGATACCGGGCAGATGCTGCAACGTTGCAGATAGCTTTAATGAACAG 720
QY 241 LeuAspValThrGlyValThrLysLysLeuAspGlyAspThrArgGlyGlyValAlaAla 260
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 721 CTGGAAATCTAGGGGTTGTTATTAACCAAAATGACGGGATACTCGTGGGGGGCTGG 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 781 CTGTCAATTCGGGCAAGTAAACGGCGCTCCGATTAATTTGCTGCTGTGGAATAATTA 840
QY 281 AspGluLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 841 ACCGATTTAGAAATTTCCATCCGATCGTATGCGATGCTATCTCTAGTATGGGGAC 900
QY 301 ValLeuSerIleGluLeuAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 901 ATGTTGACGCTAATTTGAAAGGCGCAACAGATTACGATGAGAAAAAGCAGAAAGTCT 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 961 GCTCAAAAAATGAAAGAAACAGTTTGACTTTAAGCATTTCAATGACCAATTGGATCAA 1020
QY 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1021 GTATGGCATGGCCAGTGAAGACTTATTAATAAGATCCCTGGAATGATGATACAG 1080
QY 361 LysGlyLeuAspLysLysLeuMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1081 CCTGATTTGAAATGTCAAAAGTCCATCCAAAAGATGTGGCAGAAACGGCGGATGTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1141 CTATCAATACCCCTCGACAGCGTGAATTCCTGATCTATTAATTCCTAGTCCCGCTCC 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1201 AGAATTCGACGCTGCTCAGGAATGCTGCTGTTGAAGTCATCTGATGATTAACATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThr 431
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1261 AAAGAATCCAAAAAATGATCAACAAATGTCCAAAGGGGATATGACAAATTCCTGGTATG 1320
QY 432 -----GlyGlyGlyLysLysGlyLysArgAsnGlnMetGlnAsn 446

Db 1321 GATCAATCTAGTGGCGGTTAAGGCAAGTTAGTAAATG---GCCATGAATCT 1377
QY 447 MetLeuLys 449
|||||:::|||||
Db 1378 ATGATGAAG 1386
RESULT 9
AAS53127
ID AAS53127 standard; DNA; 1434 BP.
XX
AC AAS53127;
XX
DT 13-FEB-2002 (first entry)
XX
DE Enterococcus faecalis DNA for cellular proliferation protein #555.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PE 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-255625P.
PR 22-DEC-2000; 2000US-287931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX P-PSDB; AAU35268.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS Claim 27; Seq ID No 6764; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1434 BP; 500 A; 230 C; 340 G; 364 T; 0 other;
Alignment Scores: 1.77e-119 Length: 1434
Pred. No.:

Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 23 Gaps: 2

US-09-943-108a-2 (1-455) x AAS53127 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGCGTTTGGAGCTTTAAACACCGCTTACACAGCGAATGAGTAAATCCGTCGTAAAG 60

QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAGAAGTTCCGAAGCGAGCTTAAAGAAATGATGCGAAGAAATCCGTTGCTTTATTA 120

QY 41 GluAlaAspValAsnPheLysValValLysGluPheLeuLysThrValSerGluArgAla 60
Db 121 GAAGCCGAGCTTAATTTACAAGTGGTCAAGATTTCAAAAACGCGTCAGAGAACGGGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80
Db 181 GTAGAGTCGAGTATTAGAAACCTTATCCAGCCCAACAAATGTAAATAATTTGTTGAT 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysLeuMetSerAsnLysPro 100
Db 241 GAAGAATTAACGAAACGTTAGCTTCAGAAACGGTTGAACCTGAATAAATCTCCAAAAATC 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlnAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCGACAGTGAATATGATGACAGGGTTACAGGGGCTGTAAACAACACTTTTACTGGTAAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCAAAACACTTAATGAAACACTGAACGCTCTCGCTTAAACGCTGCTGACGTT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
Db 481 TTGTATATGGGAACAGATGCTAATCCAGTGGAAATTTGTTCTCAAGGGTTAGCATTAACA 540

QY 181 LysGluGluHisLysAspPheValIleLeuAspThrAlaGlyArgLeuHisLysLeuGlu 200
Db 541 AAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysProAsnGluLysLeuMetLeu 220
Db 601 GCTTTAATGGAGCAATTAACAAATTAAGAGTGGCTAATCCCAATGAATTCGTGTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTGTGTATGCGATGACGGCGCAAGATGCTCAACGTTGCGAGATGTTTATGAAACAG 720

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGGAATTTACTGGGGTGTATTATACAAATGGACCGGACTACTCGTGGGGGGCTCG 780

QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCAATTCGGCAGTAACGGCGCTCCGATTAATTTGTCGGTCTCGTGTGAAAAATTA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACCGATTTAGAAATTTCCATCCCGATCGTATGTCGAGTCTGATCTAGTATGCGGAC 900

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGTTGAGCGCTAATTTGAAAAAGCGCAACAGATTACGATGATGAAAAAGCAGCAAGACTT 960

QY 321 GluLysLysMetArgGluSerPheThrLeuAspPhePheLeuGlnGlnLeuAspGln 340

Db 961 GCTCAAAAATGAAGAAACACAGTTTIGACTTTAACGATTTTCATTCGCAATTCGATCAA 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetLysMetLysProGlyMetAsnLysMet 360
Db 1021 GTTATGGCAGTGGGACGATTAAGACTTATTAAAAATGATCCTCGAATGAGTAACATG 1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLysPheHisLysLeuAlaIleIle 380
Db 1081 CTGTGATTTGAATGTCAAAGTCGATCCAAAAGATGGCGACGGAAACGGGCGATGGTC 1140

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CTATCAATGACCCCTGCAGAACCTGAAATCTCTGATCTATTAAATCTAGTCGCGTCGC 1200

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnLysValAsnArgLeuMetLysGlnPhe 420
Db 1201 AGAATTCAGCTGGTTCAGGAAATAGTGTGTTGAAGTCAATCGTATGATTAAACAATTT 1260

QY 421 AsnAspMetLysLysMetMetLysGlnPheThr----- 431
Db 1261 AAGAAATCCAAAAAATGATGCAACAAATGTCCAAAGGGGATATGAACATTCCTGGTATG 1320

QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMetGlnAsn 446
Db 1321 GATCAATGCTAGTGGCGCGCTTAAAGCAAGTTAGTAAATG---GCCATGAATCGT 1377

QY 447 MetLeuLys 449
Db 1378 ATGATGAAG 1386

RESULT 10
AAAX12974
ID AAX12974 standard; DNA; 6729 BP.
XX AAX12974;
XX 19-MAR-1999 (first entry)
XX Enterococcus faecalis genome contig SEQ ID NO:37.
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX Enterococcus faecalis.
XX W09850555-A2.
XX 12-NOV-1998.
XX 04-MAY-1998; 98WO-US08985.
XX 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046655.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Dillon PJ, Kunsch CA;
XX WPI; 1999-045171/04.
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
XX - used to develop products for the detection of Enterococcus and for
XX infection.
XX Claim 1; Page 406-409; 2084pp; English.
XX A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AAX12938 to AAX13919 represent these nucleotide sequences which are
XX primary nucleotide sequences, also known as contigs. The computer-based

system can identify fragments of the *Enterococcus faecalis* genome with commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, in vivo or in vitro. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.

SQ Sequence 6729 BP; 2131 A; 1181 C; 1357 G; 2053 T; 7 other;

Alignment Scores:

Pred. No.:	1.48e-118	Length:	6729
Score:	1417.50	Matches:	273
Percent Similarity:	77.5%	Conservative:	86
Best Local Similarity:	58.96%	Mismatches:	89
Query Match:	62.3%	Indels:	15
DB:	20	Gaps:	2

US-09-943-108A-2 (1-455) x AAX12974 (1-6729)

QY	1	MetAlaPheClnUcLYeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Db	3722	ATGCGCTTTTGAGAGTTTAAAAACACCGCTACAGCGCAATGAGTAATAATCCGCTGCTAAG	3781
QY	21	GlyLysLeuThrGluAlaAspLysLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	3782	GGAAAAATTTCGAGCGCGATPAAAGAAATGATCGGAGAAATCCGCTTGCGCTTAATTA	3841
QY	41	GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla	50
Db	3842	GAACCGCAGCTTAATTTACAAGTGTCTCAAGATTTCACAAAACCTGTACAGAAAGCGCA	3901
QY	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysLysValGln	80
Db	3902	GTAGAGCTCGAAGTATTAGAAACCTTTCACACCGCCACCAAAATGTGTAATAATTTGTGAT	3961
QY	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	3962	GAAGAAATTAACGAAACGTTAGGTTTCAGAAACGGTTGAACCTGAATAAATCTCCAAAAATC	4021
QY	101	ProThrValValMetMetValGlyLeuGlnGlyAlaCysThrThrAlaGlyLys	120
Db	4022	CCGACAGTGATTATGATCAGACGGTTTACAGGGCTGGTAACAACTTTTACTCGTGAAT	4081
QY	121	LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLys	140
Db	4082	TTAGCAAAACACTTAAATGAAACACTGAAACGCTCGTCCGCTTTTAATCGTGGTGACGTT	4141
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspLysProVal	160
Db	4142	TATCGTCCAGACGCGATTGATCAGTTGAAGTTTTTAGGTCAACAATAGAAGTCCCGCT	4201
QY	161	TyrSerGluClyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	4202	TTTGATATGGACACAGATGCTAATCCAGTGGAAATGTTCTGCACAGGTTAGCAATTAGCA	4261
QY	181	LysGluGluHisLeuAspPheValIleThrLeaAspThrAlaClyArgLeuHisLleAspGlu	200
Db	4262	AAAGAAAGARAATGATTATGTCTTAATTCATGACGCGCGCTTTACACATTTAGCAAG	4321
QY	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	4322	GCTTTAATGACGAATTCGAACAAATTAAGAAGTTGGCTGATCCCAATGAATTTCTGTGA	4381
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaCysLusPheAspAspGln	240
Db	4382	GTTCGTGATCCGATGACGGGCGAGATGCTGTCAACGTTGCAGATAGTTTTATATGACAC	4441
QY	241	LeuAspValThrClyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla	260

XX Berka R, Clausen IG;
 PT WPI; 2002-416684/44.
 XX Monitoring differential expression of several genes in first Bacillus
 PT cell relative to expression of same genes in one or more second
 PT Bacillus cells, by using substrate containing Bacillus genomic
 PT sequenced tag array -
 XX Claim 4; SEQ ID NO 1847; 200bp; English.
 XX The invention describes a method of monitoring differential expression of
 CC genes in a first Bacillus cell relative to expression of the genes in
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
 CC isolated from Bacillus cells to a substrate containing array of Bacillus
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first Bacillus cell relative to expression of the same genes
 CC in one or more second Bacillus cells. The method is useful for monitoring
 CC global expression of several genes from a Bacillus cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which Bacillus cells adapt to changes in culture conditions.
 CC environmental stress or other physiological provocation. Extensive
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1136 BP; 338 A; 263 C; 326 G; 209 T; 0 other;
 Alignment Scores:
 Pred. No.: 4.18e-118 Length: 1136
 Score: 1402.00 Matches: 270
 Percent Similarity: 87.30% Conservative: 60
 Best Local Similarity: 71.43% Mismatches: 48
 Query Match: 61.65% Indels: 0
 DB: 24 Gaps: 0
 US-09-943-108A-2 (1-455) x ABK74556 (1-1136)
 QY 1 MetalPheGluLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTCGAGGATTAGCCACCGCATGTGACGCAACGATCTCAAAAATCCCGGAAAA 60
 QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGAAGAGTAACGGAGCAAGACGCTCAAGAGATGATGCGGAAGTCCTGTTAGCGTCCTT 120
 QY 41 GluAlaAspValAsnPhelYsValValLysGluPhelLysThrValSerGluArgAla 60
 Db 121 GAAGCGGACGTCATATTTAAATTCGTGAAAGATTTCGTAAGAGTAAAGCGAAGCGACT 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlnGlnValLysLysLeuValGln 80
 Db 181 GTCGGACAGACGTTATGAAGACCTGACGCCGACAGACGATCAATTAAGTGTAAAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGlnAsnThrSerLysAsnMetSerAsnLysPro 100
 Db 241 GAAGAGCTGACGAGCTGATGGCGCGGAGGAAAGCAAGATTTCGCTCCAAAGGCTCG 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CGGACCGTCATCATGATGTCGTCTGCAAGGGGCGCGTAAACACGACGTCGACCGGAAAG 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLe 140

Db 361 CTGTCACACCTGCTTCGAAAAAACAATACCGCAATCGCTGCTGCGCACGACATC 420
 QY 141 TyrArgProAlaAlaLysGlnLeuGlnThrValGlyLysGlnLeuAspLeuProVal 160
 Db 421 TACCGCGCGGCGGATCAACACAGCTGGAAACGCTCGGCAAGCAGCTCGATATGCGGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
 Db 481 TTTTCACCTGGGTACAGAGCTGCTCGGTGGAGATTGCCAAGCAGCGCCATCGAAAGCG 540
 QY 181 LysGluGlnHisLeuAspPheValLysLeuLeuAspThrAlaGlyArgLeuHisLeuAspGln 200
 Db 541 AAGAGAGATCATACAGCTACAGCTTCTCATCATGACGCGGACGCCCTCATATTGATGAA 600
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220
 Db 601 GAGCTGATGGATGAGCTGAGCAGGCTAAAGAAACCCCTCAGCCGGAAGAGATTTCCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 661 GTCGTTGACTCCATGACGGGTCAAGCAGCGGTCAATGTTGCCAAAAGCTTTAACGAACAG 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 CTCGCTGTGACAGGTGTCATCTTGACGAAGCTTGACGCGCATACTAGAGCGGACGCGCC 780
 QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
 Db 781 TTGTCAATCCGCGCGCTGACAAACACGCGCATCAAAATTCGCGGAATGGCGGAAAAGCTT 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLeuGlyMetGlyAsp 300
 Db 841 GATGCGCTGGAGCGCTTTCATCTGACGCGATGGCTTCGCGGATACTCGGAAGGGCGAT 900
 QY 301 ValLeuSerLeuLysGlnAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
 Db 901 GTGCTCACCTTATCGAGAAAGCCCGACGCCAAGCTAGACGAAGAAAGCGGAGGACTC 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 Db 961 GAGCAAAAATGAAAAATATGAGCTTTAGCTGCGACGACTTTTGGAAACAGCTCGGACAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspLeuMetLysMetLysProGlyMetAsnLysMet 360
 Db 1021 GTCCCAATATGGGCGCGCTTGAGACCTGATTCAATGATGCCCGCGCAGCGCAAAATG 1080
 QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLysAspHisLysAla 378
 Db 1081 AAGGCTTAAAAAAGCTGAAGGTTGAGTAAACAGCTCAGCCATATCGAAGCG 1134
 RESULT 12
 AAH90837
 ID AAH90837 standard; DNA; 1605 BP.
 XX AC AAH90837;
 XX XX
 DT 02-OCT-2001 (first entry)
 XX XX
 DE 2CFE 28 coding sequence.
 XX XX
 KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
 KW CFE; CEG; Conserved essential gene; bacterial infection;
 KW antisense therapy; antibiotic resistance; ds.
 XX Streptococcus pneumoniae.
 PN WC200149721-A2.
 XX 12-JUL-2001.
 XX 29-DEC-2000; 2000WO-US355604.
 XX XX

```

PR 30-DEC-1999; 99US-0174089.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
XX Thanassi JA;
XX WPI; 2001-496721/54.
XX
XX Nucleic acids encoding conserved essential genes involved in bacterial
XX replication which are potential targets for the treatment of antibiotic
XX resistant bacterial infections -
XX Claim 30; Fig 49; 380pp; English.
XX
XX The present invention relates to nucleic acids (AAH90701-AAH90918)
XX encoding polypeptides (AAH01002-AAH0114), which are essential for the
XX viability of a bacterial cell wall. The acronym CFE stands for "CEG for
XX Expression", where CEG detects for "Conserved Essential Gene". The nucleic
XX acids are useful for detecting the presence of proteins essential for the
XX viability of a bacterial cell wall in samples such as cells, tissues,
XX biological fluids, blood, serum, nose, ear or throat swabs with ligands,
XX and for detecting corresponding target nucleic acid molecules with
XX complementary sequences. The nucleic acids are also useful for
XX determining whether a genomic nucleotide sequence of interest is
XX essential for viability of a bacterial cell or whether it resides within
XX an operon, by integrating an exogenous nucleotide sequence comprising a
XX portion of an open reading frame of the genomic sequence of interest
XX (comprising 200-500 base pairs) into the genomic sequence of interest
XX which confers a selectable phenotype to the cell, and determining cell
XX viability with a selection agent such as chloramphenicol. The nucleic
XX acids and proteins are also useful as vaccines and for treating bacterial
XX infections with gene therapy and antisense therapy. The nucleic acids
XX also enable identification of targets suitable for the treatment of
XX antibiotic resistant bacterial infections.
XX
XX Sequence 1605 BP; 484 A; 317 C; 393 G; 411 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 9.54e-110 Length: 1605
XX Score: 1312.00 Matches: 254
XX Percent Similarity: 74.55% Conservative: 90
XX Best Local Similarity: 55.34% Mismatches: 101
XX Query Match: 57.70% Indels: 14
XX DB: 22 Gaps: 2
XX
XX US-09-943-108A-2 (1-455) x AAH50837 (1-1605)
XX
XX Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
XX Db 1 ATGGCATTTGAAAGTTTAAACAGACGTTTCGAGACGTCGTTTAAATAATCTACGTAAAAA 60
XX
XX Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
XX Db 61 GGAAAAATCTGATCTGATGTCGACAGAGCCAAACCAAGAAATTCGCTGGCCTTGCCTC 120
XX
XX Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
XX Db 121 GAGCGCGAGCTTGCCCTGCTGTTGTAAGGACTTTATCAAGAAAGTTAGTGAGCGTGCA 180
XX
XX Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
XX Db 181 GTGGCGCATGAGGTGATGATACATTAATCTGCGCACAGATATTATAAATCGTTGAT 240
XX
XX Qy 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
XX Db 241 GAGGAAGTACAGCCGCTTTAGGTTCTGATACGGCAGAAATATCAAGTCACCTAGATT 300
XX
XX Qy 101 ProThrValMetMetValLysLeuGlnGlyLysThrThrAlaGlyLys 120
XX Db 301 CCAACCATCATGATGTTGGTTTACAAGGGCTGGTAAACACACCTTTGCTGGTAA 360
XX
XX Qy 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

```

```

Db 361 TTGCCCAACAACCTCAGAAAGAAATGCTGCTCTTGTGATGGTGGCGGGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCTGCCATTGACCACTTAAGACCTTGGGACAAACAGATTGATGCTGCTGC 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTGGCACTTGGACAGAAAGTACCAAGCTGTGAGATTGTACCTCAGGTTTGGAGCAAGCC 540
Qy 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200
Db 541 CAAACTAATCATACGACTATGCTTGATTGATACTGCGGCTCTTTGCCAGATTGATGAG 600
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220
Db 601 CTCCTCATGAATGAGCTTCGTGATGTGAAGCATTTGGCTCAACCAATGAAATCTTGCTT 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATGTGTCAGGAGACGCAATGTTGCGCGTGAGTTTATGCTCAG 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 TTGGAAGTGAAGTGGGTCATCCTTACCAAGATTGATGCGATACCTCGTGGTGGTGGT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
Db 781 CTCTCTGTTGCTGCACATCTACTGGAAACCAATCAAGTTTCACCTGATCAGGTGAAAGATT 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATTTGAAACCTTCCACCCAGACCCGATGCTACCCGATTCCTTGGTATGGGGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCATTCTTGAATGAGAAAGCTTCTCAGGAATACGATGACAAAAAGCCCTTGAATG 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATCGCGAAACACCTTGATTTTAATGATTTTCATCGATCAATTAGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCCAAATATGGGCGCATGGAAGACTTGTCTCAAGATGATTCAGATATGGCCACAAT 1080
Qy 361 LysGlyLysAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAAAACATGAAGTGGATGAACGCCAGATTGCTCGTAAACGTCGCAATTGTG 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACACCTGAGAGCGCTGAAACCCAGATTTGTTAAATCCAGCCGCTGCCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTGCTGGTCTGGAATATACATTCGTCGAAGTCAATAAATTCATCAAGGACTTT 1260
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
Db 1261 AACCGCGCTTAAACAGCTCATCGGGGTGTTATGCTGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAGCAAGCAATGGGGATTAATCCAAATAACCTTCT 1350
XX RESULT 13
XX ID AAS55519
XX XX AAS55519 standard; DNA; 1572 BP.
XX AC AAS55519;
XX XX

```

DT 13-FEB-2002 (first entry)
 XX Streptococcus pneumoniae DNA for cellular proliferation protein #90.
 DE Antisense: ds; prokaryotic cellular proliferation gene;
 XX antibiotic; antibacterial; drug design.
 KW Streptococcus pneumoniae.
 XX
 OS WO200170955-A2.
 XX
 PN 27-SEP-2001.
 PD
 XX
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253825P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI
 DR WPI; 2001-611495/70.
 DR P-PSDB; AAU37660.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX
 PS Claim 27; Seq ID No 9156; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1572 BP; 476 A; 300 C; 386 G; 410 T; 0 other;

Alignment Scores:

Pred. No.: 2.65e-109 Length: 1572
 Score: 1307.00 Matches: 253
 Percent Similarity: 74.73% Conservative: 90
 Best Local Similarity: 55.12% Mismatches: 102
 Query Match: 57.48% Indels: 14
 DB: 23 Gaps: 2

US-09-943-108A-2 (1-455) x AAS55519 (1-1572)

QY 1 MetaAlaPheGluGlyLeuSerGluArgLeuGlnAlaIleMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTTTGAAGTTTAAACAGACGCTTTGCAGACGCTTTTAAATAATCTACGTAAAAA 60
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

Db 61 GGAATAATCTCTGAATCTGATCTGATCCAGAGCAACCAAGAAATTCGCTGSCCTGCTC 120
 QY 41 GluAlaAspValAsnPheLysValLysGluPheLleLysThrValSerGluArgAla 60
 Db 121 GAGCGCAGCGTTCCTGCTGCTTTGTAAGACTTATCAAGAAAGTTTCGTGACGCGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 Db 181 GTCGGGATGAGTGCATTCATACCTTAATCCTCGCAACAGATTATTAAATGCTTGAT 240
 QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAGGAAGTACACCGCTTTAGTTCCTGATACGCGCAGAAATATCAAGTCACCTAAGATT 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCACCATCATCATGATGTTGTTTACAGGGCTGTGTAACAAACCTTTCTGGTAAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTGGCCAACTCAAGAAAGAAATCTCTCTTGTGATGTTGCGCGGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATGTCCTCAGCTGCCATTCAGCAGCTTAAGACCTTGGCAACACAGATTGATGCTGCTC 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TTTGCACCTGGACAGAAAGTACACGCTGTGAGATTGTAGCTCAAGTTTGGAGCAAGCC 540
 QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
 Db 541 CAACTAATCATCAACGACTATGCTTGATGATCTCGGGTCTGTTGAGATTGATGAG 600
 QY 201 AlaLeuMetAsnGlnLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220
 Db 601 CTCCTCAGTATGAGCTCTGATGTGAAGCATGTGCTCAACCAATGAATCTGCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
 Db 661 GTCGTTGATGCTATGATGTTGTCAGGAAGCAGCACTGTTCGGCTGAGTTTAACTGCTCAG 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 TTGGAAGTACCTGGGTCATCTTACAGATGTGCGGCTACTGCTGCTGCTGCTGCT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
 Db 781 CTGCTGTTGCTCATTACTTGGAAAAACCAATCAAGTTCACTGCTACAGTGAAAAAGATT 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 ACGGACATTTGAACCTTCCACCGACCGCATGCTACCGTATCTTGTATGGGGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
 Db 901 ATGCTCATTGTTGAGAAGGCTTCTCAGGAATACAGTACAGTACAGTACAGTACAGTAC 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 Db 961 GCTGAGAAGATGCGCGAAAAACCTTTGATTTTAAATGATTCATCGATCAATTAGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTGCAAAATATGGGCGCGATGGAAGCTTGTCTCAAGATGATTCAGGATGCGCAACAT 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeu 380
 Db 1081 CAGCCCTTCAACATGAAGGTTGATGAACCGCAGATTCGCTAAGCGTCCATGTTG 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400


```
Db 601 CTCTCATGATGAGCTTCGTGATGTGAACAATGGCTCAACCAATGAAATCTTGCTT 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTTGATGCTATGATTCAGAGAGCAACCAATGTCGCGTGAAGTTAATGCTCAG 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 TTGGAAGTGACGGGTCTATCCTTACCAAGATGATGGGATACTCTCGTGGTGGTGGT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCTGTCTGCACATTAAGTGAACCAATCAAGTTCAGTGGTACAGTGAAGAAAT 840
Qy 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATTGAACCTTCACCCAGCCGATGTCGCCGTATCCTTGGTATGGGGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACITTTGATTCAGAAAGCTTCTCAGGATTCGATGACAAAGAGCCCTTGAATG 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATCGCGGAAACACCTTTGATTTAATGATTCATCGATCAATTAGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCAAAATATGGGCCGATGGAGACTTGTCAAGATGATTCAGGTATGGCAACAAT 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAAACATGAAGTGGATGAAGCCAGATTGCTGTAAACGTGCCATTGTG 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACACCTGAAGAGCGTGAACCCAGATTGTTAAATCCCAAGCGCTGCCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGGTCTGGAAATACATTCGTAAGTCAATAAATTCATCAGGACTTT 1260
Qy 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACGAGCTAAACAGCTCATGCGGTGTTATGCTGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAATGGGATTAATCCAAATACCTTCCT 1350
```

Search completed: February 25, 2003, 01:52:44
Job time : 421 secs


```
Db 61 GGTAAACTTACGAGCTGATATAAGATAATGATGGCTGAAGTAAGATACGGTATTT 120
Qy 41 GluAlaAspValAsnPhelLysValValLysGluPheLeuLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAACCTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGGCA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
Db 181 TTAGGTTCGGATGTAATGCATCATTAACACCGGCAACAAAGTATTTAAATAGTTCAA 240
Qy 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAGTGTGAGGGTGAGAAATATACATCGATTAATATGTCAAATFAACA 300
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTATGATGGTTGGTTTACAAAGGTGCTGTAACAAACAACTCAGGTAAA 360
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATATTGATCGGTAAATATACAAACAAACCTATGTTAGTTCACGCGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCAATCAATCAACAAACAGTAGGAAACAAATTTGATATTCCTGTA 480
Qy 161 TyrSerGluLysAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGAGATCAAGTAAAGCCACAAACAAATTTGAATATGATCAATTAACATGCT 540
Qy 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAGAGAACAATTTAGACTTTGTAATCATTTGATACAGCAGTCGATACACATCGATGAA 600
Qy 201 AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGlnIleLeuMetLeu 220
Db 601 GCATTGATGACGAATTAAGAGAGTAAAGAAATTTGTAACCAACAAACGAAATTAATGTA 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTCGATCAATCAGCGGGTCAAGATGCTGTCATATTTGAGAACTTTTTCACCATCAA 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 CTTGATGTCACAGTGTACCTTAACATAATTAATAGATGGTGATACACGCTGGTGGTCAGCT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTAATCTATTCGTGCGTGACCAAAACCAATTAATTTGTTGATGATGAGTGAAGAATTA 840
Qy 281 AspGlyLeuGluPhePheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTAGACTATTCCATCTCGAAGCTATGCGCATCCGCTATTTAGGTATGGGTGAT 900
Qy 301 ValLeuSerLeuIleGluValAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAATTTGAAAAGCCCAACAGATGTCGATCAGAAAGCAAGATTTA 960
Qy 321 GluLysIleMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAGATCGGTGACTATCGCTTTACTTTAGATGATTTTTTTAGAACAACTTATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAATCTAGGACCACTGGATTAATTAAGAAATGATCCAGGATTAATGAATAATG 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLys 380
Db 1081 AAAGGGCTAGTAAGCTTAATATAGTGAAGAAAGCAAAATTTGATATTAAGGCCATATC 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
```

```
Db 1141 CAGTCATGACGGCGCTGAAGAAACAAATCCACACACATTCGATGTCACGTAAAGAG 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGCTGCTGCTGTCATTACAAAGAAAGTCAATCGTTTGATGAACAATT 1260
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
Db 1261 AACGATGAAGAAATGATGAACAATTCACCTGGTGGCGGTAAAGGTAAAGGTAAAGTAA 1320
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCATCAATGCAAAATATGTTAAAGGTATGATTTACCGTTT 1365

RESULT 2
US-09-134-001C-1958
; Sequence 1958, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1958
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1958

Alignment Scores:
Pred. No.: 2,35e-165 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: Gaps: 0

US-09-943-108A-2 (1-455) x US-09-134-001C-1958 (1-1014)
Qy 1 MetalApheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGGCAATTTGAGGATATTCGGATCGCTTACAGCCAGCATGCAAAAATCGGTGTAA 75
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAGAAGTAAACAGACAGATATTAAACGATGATGCGTGAAGTACGATTACGCTTATG 135
Qy 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 136 GAAGCCGATGTACTTCAAGTTGTTAAGCAATTTGTAAGATGTTTCAGACGAGCG 195
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
Db 196 CTAGGTTCTGATGTCATGCAATCTTTAAACCTCGCCACACAGGTATTAAATCGTACAA 255
Qy 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256 GAAGAACTTCTAGTTATATGGGTGAGAAATATCTTCCATTAAGATGCAACAAACCA 315
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 316 CCAACTGTTGTCATGATGTCGTCGCTTACAGTCAGGTAAACGACGACGACGAGTAAG 375
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 376 TTGGCATTAATTAATCGTAAAAAATATAACAAAAAACCTTTACTTGTGCGCAGAGATATT 435
```



```

|||||:::|| ::|| | ||||| ||||| ||| |||||:::
Db   2253 CTGTCGTGGTCCACATCTAGAAACCAACTCAAGTTCACTGGTGACAGTGGAAGAATT 2194

QY   281 aspGlyLeuLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   2193 ACGGACATTGAACAACCTCCACCAGACGCCGATGCTAGCCGTATCCTTGTAAGGGGAT 2134

QY   301 ValLeuSerLeuIleGluLysAlaGlnGlnaspValasplngLuLysAlaLysaspLeu 320
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   2133 ATGCCTCACTTTGATTGAGAAGCTTCTCAGGAATTCAGCATGAAACAAAAGCCCTTGAATG 2074

QY   321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuaspGln 340
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   2073 GCTGAGAGATGCGGAAACACCTTGAATTATGATGTTCTCGATCAATTAGATCAG 2014

QY   341 VallhsAsnLeucyProLeuAspAspileMetLysMetileProGlyMethenLysMet 360
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   2013 GTGCAAATAATGGCCGCGTAGAAGACTTGCTCAAGATGATTCAGGTATGGCCAACAT 1954

QY   361 LysGlyLeuAspLysLeuasnMetSergLnuLysGlnLasphIsLelyslarlle 380
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   1953 CGAGCCCCTCAAAACATCAAGGTGGATGAACGCCAGATTCTCGPAAACGTGCCATTGTG 1894

QY   381 GlnSerMetThrProAlaGluArgsnAsnProAspPhrLeuAsnValSerArgLysLys 400
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   1893 TCTTCGATGACACCTGAAGAGCGTGAAAAACCCAGATTGTGTTAAATCCAGCCGTCGCCGT 1834

QY   401 ArgIlealaLysGlySerGlyArgserLeuglndluValasArgLeuMetLysGlnPhe 420
      :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   1833 GGTATTCTGCTGGTGTCTGGAAATACATTCTCGAAGTCAATAATTCATCAAGGACTTT 1774

QY   421 AsnAspMetLysLysMetMetLysGlnPheThrGlvgLyGlyLysGlyLysLys 440
      :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db   1773 AACAGCGTPAACACTCATGAGGCTGTATGTCGGG----- 1735

QY   441 ArgasnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeupro 454
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1734 -----GATATGAATAAAATGATGAAGCAATGGGGATTAAATCCAAATAACCTTCT 1684

RESULT 4
US-09-035-382-3
; Sequence 3, Application US/09035382
; Patent No. 6284515
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES
; FILE REFERENCES: GM50035
; CURRENT APPLICATION NUMBER: US/09/035,382
; CURRENT FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: 60/057,890
; EARLIER FILING DATE: 1997-09-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-035-382-3
```

Qy	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgLeuValArgLeuAlaLeuPhe	40
Db	61	GGTAAACTTACTGAAGCTGATATAGATAAATGCTGCTGAAGTAAAGATTAGCGTTATT	120
Qy	41	GluAlaAspValAsnPheLysValValLysGluPheLileLysThrValSerCLeuArgAla	60
Db	121	GAGGCTGACGTAACTTAAAGTGGTAAGAAGATTTATTAACACGTATCAGACGGCCA	180
Qy	61	LeuGlySerAspValMetIleSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
Db	181	TTAGGTTCCGATCTAATSCAATCATTAACACAGGCGCAACAAGTTATTAATAATAGTTCAA	240
Qy	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	241	GATGAATTAACGSAAGTTGATGGGTGGAGAAATACATCGATTAATATGTCAAAATAACCA	300
Qy	101	ProThrValValMetMetValClyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Db	301	CCCTACTGTTGTTATGATGGTGTTCACAGGTGCTGGTAAACAACAACATCGCAGGTAA	360
Qy	121	LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	361	TTAGCATTTATGTGCGTAAAAAATACACAAAAAACCTATGTAGTTCCACAGATATT	420
Qy	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	421	TATCTGCCAGCAGGATACATCAATACAAAACAGTAGGGAACAATGTATATCTCTGTA	480
Qy	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	481	TACACTGAAGGAGATCAAGTAAGCCACAAACAAATGTAACTATGCAATAAACATGCT	540
Qy	181	LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	541	AAAGAGAACAACTTACACTTGTATCTTCATCAGCAGGTGATTCACATCGATGAA	600
Qy	201	AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	601	GCATTTGATGACAGCAATTAAGAGAAAGTAAAGAAATGTCTAAACCAACGAATATGTGA	660
Qy	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
Db	661	GTGTGCGATTCAATGACGGGTCAAGATGCTGTCTATGTTCAGAACTCTTTTCACGATCAA	720
Qy	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla	260
Db	721	CTTGATGTCACAGGTGTTACCTTAACTAAATTAGATGGTGATACCCGTGTGTCAGCT	780
Qy	261	LeuSerIleArg 264	
Db	781	TTATCTATCTGT 792	
		RESULT 5	
		US-08-923-772-1	
		; Sequence 1, Application US/08923772	
		; Patent No. 5972651	
		; GENERAL INFORMATION:	
		; APPLICANT: Black, Michael T.	
		; TITLE OF INVENTION: NOVEL IFH	
		; NUMBER OF SEQUENCES: 2	
		; CORRESPONDENCE ADDRESS:	
		; ADDRESS: Dechert Price & Rhoads	
		; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre	
		; CITY: Philadelphia	
		; STATE: PA	
		; COUNTRY: US	
		; ZIP: 19103	
		; COMPUTER READABLE FORM:	
		; MEDIUM TYPE: Diskette	
		; COMPUTER: IBM compatible	
		; OPERATING SYSTEM: DOS	
		; SOFTWARE: FastSeq for Windows Version 2.0	

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,772
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dickinson, Todd Q
 ; REGISTRATION NUMBER: 28,354
 ; REFERENCE/DOCKET NUMBER: GML0080
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-994-2252
 ; TELEFAX: 215-994-2252
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1569 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-923-772-1

Alignment Scores:
 Pred. No.: 4,49e-138 Length: 1569
 Score: 1297.00 Matches: 251
 Percent Similarity: 74.51% Conservative: 91
 Best Local Similarity: 54.68% Mismatches: 103
 Query Match: 57.04% Indels: 14
 DB: 2 Gaps: 2

US-09-943-108a-2 (1-455) x US-08-923-772-1 (1-1569)

QY 1 MetAlapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTTGAAGTTTAAACAGAACGTTTGAGAACGCTTTTAAATAATCTACGTAATAAAA 60
 QY 21 GlyLysLeuThrGluAlaAspLysLysLysLysLysLysLysLysLysLysLysLysLys 40
 Db 61 GGAATAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 QY 41 GluAlaAspValAsnPhelLysValLysGluPheLysLysLysLysLysLysLysLysLys 60
 Db 121 GAGCGCGAGTGTGCTTGGCTTGTGTAAGGACCTTATCAAGAAAGTGTGTCGAGCGTGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysLysLysLys 80
 Db 181 GTCGGGCATGAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysAsnMetSerAsnLysPro 100
 Db 241 GAGGACCTGACAGCGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCACAT 360
 QY 121 LeuAlaLeuMetArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 140
 Db 361 TTGGCCACAACTCAGAAAGAGAAAGAAATGCTGCTCTTGTGATGATGATGATGATGATGAT 420
 QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysLysLysLysLysLys 160
 Db 421 TATCGTCAGCTGCCATGACAGCTTAAGACCTTGGGACACAGATGATGATGATGATGATGAT 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysLysLysLysLysLysLysLysLys 180
 Db 481 TTTGCACTTGGACAGAGTACCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 540
 QY 181 LysGluGluHisLeuAspPheValLysLysLysLysLysLysLysLysLysLysLysLysLys 200
 Db 541 CAAACAT 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysLysLysLysLysLysLysLys 220
 Db 601 CTCCTCATGAATGAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 661 GTCGTTGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 TTGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLys 280
 Db 781 CTGCTCTTCTGTCATCATCACTGGAACCAATCAAGTTCACGTGACAGGTGACAGGTGACAG 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLysLeuGlyMetGlyAsp 300
 Db 841 ACAGATATCGAAACCTTCCACCAGACCGTATGCTAGCCGATGCTTGGCATGGGGAT 900
 QY 301 ValLeuSerLeuLysGlnLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 ATGCTCACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
 Db 961 GCTGAGAGATGCGGAAACACCTTTGATTTTANTGATTTCTGATGATTTCTGATGATTTCTG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspLysMetLysMetLysMetLysMetLysMetLysMet 360
 Db 1021 GTGCATAAATATGGGCGGATGGAAGCTTGTCTCAAGATGATGATGATGATGATGATGAT 1080
 QY 361 LysGlyLeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 380
 Db 1081 CCACATCTCAAAACATCAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 TCTTCGATGACATCTGAAGACGTGAAGACCCAGATTTGTTAAATCCAAGCCGTCGCGGT 1200
 QY 401 ArgLeuAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 QY 421 AsnAspMetLysLysMetLysLysLysLysLysLysLysLysLysLysLysLysLysLys 440
 Db 1261 ACCAGGCTAAACAGCTCAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1299
 QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMet-----AsnLeuPro 454
 Db 1300 -----GATATGAATAAATGATGAAGCAATGGGATTAATCCAATAAAGCTTCTCT 1350

RESULT 6

US-09-385-287-1
 ; Sequence 1, Application US/09385287
 ; Patent No. 6350857
 ; GENERAL INFORMATION:
 ; APPLICANT: Black, Michael T.
 ; TITLE OF INVENTION: NOVEL fih
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dechert Price & Rhoads
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows version 2.0
 ; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-385-287-1
Alignment Scores:
Pred. No.: 4,49e-138 Length: 1569
Score: 1297.00 Matches: 251
Percent Similarity: 74.51% Conservative: 91
Best Local Similarity: 54.68% Mismatches: 103
Query Match: 57.04% Indels: 14
DB: 4 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-385-287-1 (1-1569)
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGCGATTGGAAGTTTAAACAGACGTTTCGAGAGCGCTTTAAAAATCTACGTAAAAA 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GAAAAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 120
QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60
DB 121 GAGCGGACGTTGCGCTTCCTCTGTTAAAGCACTTATCAAGAAGTTCGTGAGCGTGA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DB 181 GTGGGCGATGAGTCAATGATACATTAATCTGCGCACACAGATATATAAATCGGTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DB 241 GAGGAATGACAGCGGTTTGGTCTGATACGCGCAGAAATATCAAGTCACCTAAGATT 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CCAACCATCATCATGTTGGTTTCAAGGGCGTGGTAAACAAACCTTTGCTGGTAAA 360
QY 121 LeuAlaLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIle 140
DB 361 TTGGCCACAAATCAAGAGAGAGAAATGCTGTCCTTTGATGATGCGGCGGTATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
DB 421 TATGCTCCAGCTGCCATTGACCACTTAAGACCTTGGGACACACAGATGATGTCCTGTC 480
QY 161 TyrSerGluGlyAspGluValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
DB 481 TTTCGACTTGGACACAGATACACGCTGTTGATGTTGATGTTGAGGTTGGAGCAGCC 540
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
DB 541 CAACACTAATCAACAGACTATGCTTGTGATGATGATGCTGCGGGCTGTTGCAGATTGATG 600

```

```

QY 201 AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGluIleMetLeu 220
DB 601 CTCTCTCATGATGAGCTTCGTGATGGAAGATATTGGCTCAACCAATGAATCTGCTT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
DB 661 GTCGTTGATGCTATGATTGCTCAGGAAGCAGCAATGTTGGCGTGAGTTAATCTGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAla 260
DB 721 TTGGAAGTGAAGTGGGGTCATCCTTACCAAGATTGATGCTATCTGCTGCTGCTCTCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
DB 781 CTGCTGTTGCTCACATCACTGGAAAACCAATCAAGTTCACTGGTACAGTGAAAAATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetalaserArgIleLeuGlyMetGlyAsp 300
DB 841 ACAGATATCGAAACCTTCCACCAGACCGTATGCTAGCCGTATCTTGGCATGGGGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
DB 901 ATGCTCACTTTGATTGAGAAGCTTCTCAGGAATACGATGATCCAGTATGCGCAACAAT 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
DB 961 GCTGAGAAGATCGCGGAAAACACCTTGTGATTTTATGATTTTCATCGATCAATAGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
DB 1021 GTGCAAAATATGGGCGGATGGAAGACTTGTCAAGATGATTCAGGTATGCGCAACAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
DB 1081 CCAGCACTTCAAAACATCAAGGTGGATGAACCCAGATGCTGTAACGTGCCATTGTG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
DB 1141 TCTGATGATGACATCTGAAGAACCTGGAACCCAGATTTGTAAATCCAAAGCGTCCCGT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
DB 1201 CPTATTGCTGCTGTTGTTGGAATACATTCGTCGAAGTCAATTAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
DB 1261 AACAGGCTAACAGCTCATGCGAGGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
DB 1300 -----GATATGAATAAATGATGAAGCAATGGGATTAATCCAAATAACCTCTCT 1350

```

RESULT 7

```

US-09-221-017B-928/c
; Sequence 928, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Pp1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Pp1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Pp2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: Pp2/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 928:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...1098

US-09-221-017B-928

Alignment Scores:
Pred. No.: 6,81e-105 Length: 1098
Score: 1003.50 Matches: 190
Percent Similarity: 73.17% Conservative: 86
Best Local Similarity: 51.49% Mismatches: 96
Query Match: 44.13% Indels: 3
DB: 4 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-221-017B-928 (1-1098)

Qy 17 MetArgGlyLysGlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArg 36
Db 1098 CTCRAAGGTGAGGGGAGGAATACCGAGATCAATGTAGCCGAACCTCTCAAGATGTAGT 1039
Qy 37 LeuAlaLeuPheGluAlaAspValAlaPheLysValLysGluPheLysThrVal 56
Db 1038 CGTGCCTTTTGGATGCGGCGGTAATGTCTGTGCGGAGGAGGATTTACGGACTGTGGT 979
Qy 57 SerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIle 76
Db 978 AAGGAAAGACCATCGGCCAAACGGTGTACCTCCGTCGTCGGCGGAGATGTAGTGGT 919
Qy 77 LysIleValGlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMet 96
Db 918 AAGATGTTTCATGACGAATGGCTGCTGATGGCGGACATTCCTGTGGATATTATATCTC 859
Qy 97 SerAsnLysProProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThr 116
Db 858 AAAGGCAGT---CCGGCAGTAATCCTGATGTCGGATTGCAAGTTCGGGTGAAGACACC 802
Qy 117 ThrAlaGlyLysLeuAlaLeuMetArgLysLysTyraAsnLysLysProMetLeuVal 136
Db 801 TTCTCCGCAACATGGCCATATATGCTCAAGTCCAGAGGGCAACGACCTCTGTCTGTT 742

Qy 137 AlaAlaAspIleTyArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIle 156
Db 741 GCCTGCAGCGTGTATCGCCTCGGCTATCCAGCAGCTGATGCTCGCGCAGCAGCTC 682
Qy 157 AspIleProValTySerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAla 176
Db 681 GGGGTACCCGCTCTACAGGACCGGACAGCAAAACCGGTCGAGATAGCCCGACAGCC 622
Qy 177 LeuLysHisAlaLysGluLysGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeu 196
Db 621 ATACCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 562
Qy 197 HisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsn 216
Db 561 GCCATCGACGAGAATGATCGCGGATAGAGGCCATCAAGGCTGCCATCCAGCCGAC 502
Qy 217 GluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSer 236
Db 501 GAAGTCTCTTTTCGTGGTGGACTCGATGACCGGTGAGGACGCGGTGATACAGCAAGGAG 442
Qy 237 PheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg 256
Db 441 TTCAATACACGCTTGGATTGCGACGGTGTCTCTTACCAAGCTCGATGAGATACCCGG 382
Qy 257 GlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValIleYmet 276
Db 381 GCGGTGCGCCCTCTCGATCGTTCGTGCTGTAACAACAGCCCATCAAGTTCGTGCTAG 322
Qy 277 SerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeu 296
Db 321 GGGGAGAGATGGAACAATCGATGCTTTTCAACCGGAGCGTATGGCGACCGTATCCTT 262
Qy 297 GlyMetGlyAspValLeuSerLeuIleGlyAlaGlnGlnAspValAspGlnGluLys 316
Db 261 GGCATGGGGGACATCGTATCGCTTGGGCGGTCTCAGGAGCAATACAGACAGCGGAA 202
Qy 317 AlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGlu 336
Db 201 GCTCGCAACTGGGAGAGAAATAGCCAAAGATCAGTTCGACTTCAATGACTTCTTGC 142
Qy 337 GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGly 356
Db 141 CAGATACATCAGATCAAAAGATGGTAACTCAAGAGCTTGCATCCATGATCCCGGA 82
Qy 357 MetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIle 376
Db 81 GTGGA-----AGGCCATCAGGATATCCAGATACAGACGATCCCTTCAGAGCATC 28
Qy 377 LysAlaIleIleGlnSerMetThrPro 385
Db 27 GAGGTATCATATATTCATGAGGCC 1
RESULT 8
US-09-066-047-6
; Sequence 6, Application US/09066047A
; Patent No. 6306394
; GENERAL INFORMATION:
; APPLICANT: MURPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
; USE OF GRANULOCYTIC ERLICHIA
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

Alignment Scores:	
Pred. No.:	1.24e-57
Score:	52.50
Percent Similarity:	50.11%
Best Local Similarity:	33.84%
Query Match:	26.06%
DB:	2
Length:	2166
Matches:	156
Conservative:	75
Mismatches:	211
Indels:	19
Gaps:	9

US-09-943-108A-2 (1-455) x US-08-317-401E-3 (1-2166)

QY	1	Met	AlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
Db	86	ATG	TCTCCTCAGGATCTGGCGCGCAATCAACGCGCGCTCAATGACTGACTCGCTCC	145
QY	21	Gly	LysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	146	ACA	ATTGGACGAGAGCCTTGAGACATGATTAAAGAGATCGCGCCGCTCTG	205
QY	41	Glu	AlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla	60
Db	206	TCC	GCGAGCTCAAGTCGCGCTGGTCCAGTCCCTCCGCAAGTCCATCAATCAGCGTC	265
QY	61	Leu	GlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln	80
Db	266	AACT	TTCCTCTCCTCCCTCGCGCGTGAACAAGAGCGTTGATTCAAAGAGCGCTTTC	325
QY	81	Asp	GluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	326	GAT	GAGCTGGTTTCCCTCGTTCATCCCATCGGAGAGCCCTCCGCCCCCAAGAAGGCGC	385
QY	101	Pro	ThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
Db	386	TCC	AAGTGTATCATGTTCGTGGTCTGCAAGGTGCGGTAACACCACTGTGTACCAAG	445
QY	121	Leu	AlaLeuLeuMetArgLysLysTyrAsnLys-----LysProMetLeuValAla	137
Db	446	CTGGC	-----CGCCATACAGATCGCGGCTTCAAGACTGCCTCGTGTGT	493
QY	138	Ala	AspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyGlnIleAsp	157
Db	494	GCG	ATACCTTCGTCGTGGTTCCTTCGACAGCTGAAGCAGAATGCCACCAAGCAAG	553
QY	158	Ile	ProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeu	177
Db	554	ATCC	CTACTACGTTAGCTTGACCCARACCGACCGCCGCGCATTTGTACACCGAGGGTGTG	613
QY	178	Lys	HisAlaLysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyLeuHis	197
Db	614	GCA	AGTTCAGAAGAGCGGTTTCGAATCATCATCTGTCGATACCATGGTGTCACCAAG	673

```

Qy 198 lileaspGluAlaLeuMetAsnGluLeuLysGluValLysGluAlaLysProAsnGlu 217
Db 674 CAGGAAGAAGAGCTTTTCCACCGAAATGACCGAGATATCAGACCGCGCTCACCCCGCAGC 733
Qy 218 lileMetLeuValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPhe 237
Db 734 ACCATCTGCTGCTCGACGAGCACACFCGGTCAAGCTGCGAGTGGCGAAGCCACTCTGCTTC 793
Qy 238 AspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGly 257
Db 794 AAGCCACCGCAGACTCTGGAGCCATCATCATCACCAAGACGGATGTCACGCCGCGAGST 853
Qy 258 GlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSer 277
Db 854 GGTGTGCTATTTCGCGCGTCGCGCACACACATCCCATATTACCTCGGTGACCGGT 913
Qy 278 GlnLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGly 297
Db 914 GAGCACTCATGATCTGGAGACCTTTGAGCGGAGGCGCTTCATCCAGAGACTCTGCGGT 973
Qy 298 MetGlyAspValLeuSerLeuIleGlnLysAlaGlnGln---AspValAspGlnGlnLys 316
Db 974 ATGGCGCATATGCTGCTGGTAGAGCAGCTACAAAGCGGTGACCAAGGACTCTGCTCC 1033
Qy 317 AlaLysAspLeuGlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlu 336
Db 1034 GCCAAGAAACCTACAAGCACATCTCCGAAGGTATCTACACGCTCGGTGACTCCGCGAG 1093
Qy 337 GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGly 356
Db 1094 AACATCACTTCATCATCAAGATGGAGCCCTCTCCAGCTCTCGCGCATGATCCCGGT 1153
Qy 357 Met---AsnLysMetLysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHis 375
Db 1154 CTATCCAACTGACCGCGGGTCTGATGACGAAGACGGCTCCATGAAG--CTCCGTCGC 1210
Qy 376 lileLysAlaIleIleGlnSerMetThrProAlaGluArgAsnProAspThrLeuAsn 395
Db 1211 ATGATCTACATTTCTCGACATGATGAGCGCGCGGCGGATCTGATGGC---GATGGCAGAAAT 1267
Qy 396 Val-----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
Db 1268 TTCGTCGAACACCCAGCCGGATGGTCCGATGCTGCGGAAGCGGTACACCGCTCCG 1327
Qy 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetMetLysGlnPheThr 431
Db 1328 GAAGTGAAGACGCTGCTCCACGACCCGATGATGCGCGCATGGCCGAGTGGCCAAAGCGTGC--- 1384
Qy 432 GlyGlyGlyLysGlyLysGlyLysArgGlnMetGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1385 ---GGTGGACGAGACGACAGATGCAGCGC-----CCCAAGACATGCTCAAGGCGGT 1435
Qy 452 Asn 452
Db 1436 AAC 1438
RESULT 10
US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Iyer, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561 No. 5922561disk of No. 5922561th America, I
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```

```

; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,401E
; FILING DATE: 03-October-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4248,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (126..203, 253..1776)
US-08-317-401E-1

Alignment Scores:
Pred. No.: 2,8e-55 Length: 2877
Score: 573.50 Matches: 149
Percent Similarity: 51.16% Conservative: 71
Best Local Similarity: 34.65% Mismatches: 193
Query Match: 25.22% Indels: 17
Db: 2 Gaps: 8

US-09-943-108A-2 (1-455) x US-08-317-401E-1 (1-2877)

QY 31 MetMetArgGluValArgLeuAlaLeuPheGluAlaAspValMetLysValValLys 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 ATGATCAAGAGATCTGGCGCCCTTGCTGCGCGAGCTCAAGCTCGCGCTGGTCCAG 327

QY 51 GluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThr 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 TCCCTCGCAATCCATCAGTCCAGCGTCAACTTTCCTCTCTCTCTCTCTCTCTCT 387

QY 71 ProGlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeuMetGlyGlu 90
::: ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 388 AAGAAGCGTTTGATTTCAGAAAGCGCTTCGATGAGCTGGTTTCGCTGGTATCCCAT 447

QY 91 AsnThrSerIleAsnMetSerAsnLysProProThrValValMetMetValGlyLeuGln 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 448 CGGAGGCATTCGGTCCCAAGAGCGCGTCCCAAGCTGATCATCTTCGCTGCTACAG 507

QY 111 GlyAlaGlyLysThrThrThalacGlyLysLeuAlaLeuMetArgLysLysThrAsn 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 508 GGTCCGGTAAACACCAACCTGTGACCAAGCTGGC-----CGCCACTATCAG 555

QY 131 Lys-----LysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsn 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 556 ATGCGCGGTTCAAGACCGCCCTGTGTGCGGATACCTCCGAGCTGGTGTTCGAC 615

QY 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnVal 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 616 CAGCTGAACAGAAAGCCACCAAGCCAGATCCCTACTACGTAGCTGACGCAACAC 675

QY 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluAlaHisLeuAspPhe 187
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 676 GACCCCGCCCATCGTGGCGCGGAGGTGTGCCCAAGTTCACAGAGAGCGTTCTTCGAAATC 735

QY 188 ValIleIleAspThrAlaGlyLeuHisIleAspGluAlaLeuMetAsnGluLeuLys 207

```

```

; ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 736 ATCATTTGTGATACCACTGGTGGTCGTCATACCAAGAGAGAGAGCTCTTCACGAAATGACC 795

QY 208 GlnValLysGluIleAlaLysProAsnGlnIleMetLeuValValAspSerMetThrGly 227
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 796 CAGATTCAGACCGCGTCAACCGGACAGACCATCTCGTCTCGACAGACCATCGGT 855

QY 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247
||| ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 CAGCTGGCGAAGCCAGTCTCGCTTCAGGCCACCGAGACTTCGGAGCCATCATC 915

QY 248 LeuThrLysLeuAspGlyAspThrArgGlyGlyAlaLeuSerIleArgSerValThr 267
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 916 ATCACCAGACGATGGTCACGCGCAGGTGGTGTCTATTTCGCGCTCGCCGACCA 975

QY 268 GlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHis 287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 976 CACACTCCCATTTACTCGGTACCGGTGAGCATTGATGGACCTGGAACGTTTCGAS 1035

QY 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuGluLys 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1036 CCCAAGCCTTATCCAGAGCTCTCGGTATGGGTATATGGCGGACTGGTAGAGCAC 1095

QY 308 AlaGlnGln---AspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGlu 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1096 GTACAAGCGGTGACCAAGGACTCGGCTCGGCCAAGAAACATACAAGCATCTCAGAA 1155

QY 327 SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGlyPro 346
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1156 GGTATCTACACGCTCGCGACTTCGCGGAGAAATCATCTCCATCATGATGGGTCT 1215

QY 347 LeuAspAspIleMetLysMetIleProGlyMet---AsnLysMetLysGlyLeuAspLys 365
||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1216 CTCTCCAAAGCTCTCGCGCATGATTCGCGGTCTCTCCAACTGACCGCGGACTGTGATGAC 1275

QY 366 LeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIleGlnSerMetThrPro 385
::: ||| ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1276 GAAGACGGCTCCATGAAG---CTCGCGCGCATGATCATCATTTTCGACAGCATGACGGCC 1332

QY 386 AlaGluArgAsnAsnProAspThrLeuAsnVal-----SerArgLysLysArgIle 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1333 GCGGAAGTCCGCGGCGGCGGCAAGATGTCGTCGAACAGCCTGACCGCATGGTCCGGATC 1392

QY 403 AlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsnAsp 422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1393 GCTTGGGAGGCGGTACCAACCGTCCGGAAGTCGAAGACCTGCTCTCCAGCACCCCATG 1452

QY 423 MetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysArgAsn 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1453 ATGGCTGGTATGGCAAGCGGTC-----GGCGGACAGAAGAAGACGATCGACGT--- 1503

QY 443 GlnMetGlnAsnMetLeuLysGlyMetAsn 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1504 ---GCCAGAGACATGCTCAGGGCGGTAAAC 1530

```

RESULT 11

```

US-08-981-527A-20
; Sequence 20, Application US/08981527A
; Patent No. 6410262
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Kerkman, Richard
; APPLICANT: Broekhuizen, Cornelis P.
; TITLE OF INVENTION: No. 6410262el Secretion Factors for
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes Encoding Them and Methods
; TITLE OF INVENTION: of Using It
; FILE REFERENCE: GCX322-US
; CURRENT APPLICATION NUMBER: US/08/981,527A
; CURRENT FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: PCT/NL96/00278
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
/ SEQ ID NO 20
; LENGTH: 4370
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-08-981-527A-20

Alignment Scores:
Pred. No.: 6,96e-39 Length: 4370
Score: 432.00 Matches: 114
Percent Similarity: 53.25% Conservative: 66
Best Local Similarity: 33.73% Mismatches: 120
Query Match: 19.00% Indels: 38
DB: 4 Gaps: 10

US-09-943-108A-2 (1-455) x US-08-981-527A-20 (1-4370)

QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 2351 GATTCGGTATCTGAAAGTTTAAAGTATGCGCTTGAAGAAACAAAGAACCTCTTCAAAAC 2410
QY 24 ThrGluAlaAspIleLysIleMetMetArgGlnValArgLeuAlaLeuPheGlu----- 41
Db 2411 AAAGTGAATGATCTGTATCCCGTTACCGTAAAGTGGATGAGGATTCCTCGAAGAGCTT 2470
QY 42 -----AlaAspValAsnPhelLysValLysValLysGluPheIle----- 53
Db 2471 GAAGAGGTTCTATACCGCGGATGTCGGTTTACACCGGTTATGGAATTAATAGATGAG 2530
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 2531 CTGAAAGAAAGTCAACACGAGA-----AATATTCAAGAT 2566
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 2567 CCAGAGGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2626
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProProThrValValMetMetValGly 108
Db 2627 GAGCAAAATTCAGAACTGAACATCCAGGATGGCGCTTAAACGTAATCTCTGTGTAGT 2686
QY 109 LeuGlnGlnAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLys 128
Db 2687 GTAACCGCGCGCGGAAACAAACAGATCGGAAAGCTGTCTATAAATG---AAACAA 2743
QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsnGln 148
Db 2744 GAAGGAAATCTGTGTACTTGCCTGCGGAGACATTTAGACGCGGAGCCATTGAACAG 2803
QY 149 LeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnValLys 168
Db 2804 CTGGAGTATGGGAGAGCGCTACAGAGTGCCTGTCTAATAGCAGACGCGGAGAGCGGAT 2863
QY 169 ProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlnGluHisLeuAspPheVal 188
Db 2864 CCGCGCGCTGTCTACAGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2923
QY 189 IleIleAspThrAlaGlyArgLeuHisIleAspGlnAlaLeuMetAsnGlnLysGln 208
Db 2924 ATTTGTGATGAGCGGAGCGGCTCTCCAAACAAAGTAAATCTCAAGAAAGAGCTTGAAGA 2983
QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
Db 2984 GTAAACGTTGTTATCGAAGAGAGAGTCTCGAAGCTCGCGATGAGGTGCTGCTGCCCTT 3043
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 3044 GATGCCAGCAGCGGCGCAAAATGCAATGCTCAGGCAAAAGAAATCTCTAAAGCAACAAAT 3103
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 3104 GTTACCGGATTCCTTAAAGAGAGCTTACCGGTACCGGAAAGAGCGGTATCGTCTCTGCG 3163
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
```

```
Db 3164 ATTCCGACAGAGCTTCACATCCCGTTAACTAGTCGGTTAGAGAAATAATTATGAC 3223
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 3224 CTTACAGAAATTGATCCAGAA-----TCTATGTACGGACTC-----TTTTCA 3268
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 3269 GATTAGTGGAAAGGCCGAC-----GATTAGAAAGAGGCCCAACATC 3313
RESULT 12
US-08-981-527A-7
; Sequence 7, Application US/08981527A
; Patent No. 6410262
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Kerkman, Richard
; APPLICANT: Broekhuizen, Cornelis P.
; TITLE OF INVENTION: No. 6410262el Secretion Factors for
; TITLE OF INVENTION: Gram-Positive Microorganisms Genes Encoding Them and Methods
; TITLE OF INVENTION: of Using It
; FILE REFERENCE: GCX322-US
; CURRENT APPLICATION NUMBER: US/08/981,527A
; CURRENT FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: PCT/NL96/00278
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1015
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)...(1008)
; OTHER INFORMATION: protein secretion chaperrone
US-08-981-527A-7

Alignment Scores:
Pred. No.: 1,22e-39 Length: 1015
Score: 430.00 Matches: 114
Percent Similarity: 53.29% Conservative: 64
Best Local Similarity: 34.13% Mismatches: 118
Query Match: 18.91% Indels: 38
DB: 4 Gaps: 10

US-09-943-108A-2 (1-455) x US-08-981-527A-7 (1-1015)

QY 4 GluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeu 23
Db 64 GATTCGGTATCTGAAAGTTTAAAGTATGCGCTTGAAGAAACAAAGAACCTCTTCAAAAC 123
QY 24 ThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGlu----- 41
Db 124 AAAGTGAATGATCTGTATCCCGTTACCGTAAAGTGGATGAGGATTCCTCGAAGAGCTT 183
QY 42 -----AlaAspValAsnPhelLysValLysValLysGluPheIle----- 53
Db 184 GAAGAGGTTCTTATCAGCGGAGTTCGGTTTACACCGCTTATGGAATTAATAGATGAG 243
QY 54 -----LysThrValSerGluArgAlaLeuGlySerAspValMetGlnSerLeuThrPro 71
Db 244 CTGCAAAAGAGTCAACCCAGA-----AATATTCAAGAT 279
QY 72 GlyGlnGlnValIleLysIleValGlnAspGluLeuThrLysLeu-----MetGlyGly 89
Db 280 CCAAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 339
QY 90 GluAsnThrSer-----IleAsnMetSerAsnLysProProThrValValMetMetValGly 108
Db 340 GAGCAAAATTCAGAACTGAACATCCAGGATGGCGGCTTTAAACGTAATCTCTGTGTAGT 399
```

```
QY 109 LeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLys 128
Db 400 GTAAAGCGGCTGGGAAACAAACACGATCGGAAGCTTGCTCATAAATG---AAACAA 456

QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsnGln 148
Db 457 GAAGAAATCTGTGTACTTGCCTGGAGAGACACTTTAGAGCGGAGCCATGAAACAG 516

QY 149 LeuGlnThrValGlyLysGlnAspIleProValTyrSerGluGlyAspGlnValLys 168
Db 517 CTGGAAGTATGGGAGAGCGTACAGGAGTGCCTGTCTATTACGACAGCGGACGAGCGAT 576

QY 169 ProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPheVal 188
Db 577 CCGCGCGCTGTCTATCGACGCTGTTCATGCTCGCAAGCAAGAAATGCCGATGTATTA 636

QY 189 IleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGlu 208
Db 637 ATTGTGATACGGCAGGCGTCTCCAAACAAAGTAATCTCATGAAGAGCTTGAAAAA 696

QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
Db 697 GTAACAGCTGTATCGAAAGAGAGTTCCTGAAGCTCCGATGAGTCTGCTTGCCCTT 756

QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
Db 757 GATGCCACACCGCGGCAAAATGCAATGGCTCAGGCAAAAGAAATCTCTAAGCAACAAT 816

QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 817 GTTACCGGCAATGCTTTACGAGACTTGACGGTACGGCAAAAGCGGTATGCTTCCTT 876

QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 877 ATTCCGACAGAGCTTCACATCCCGTTAAACTAGTCGGTTTAGAGAAAAAGTTGATGC 936

QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 937 CTCGGAATTTGATCCAGAA-----TCTATGTGTACGGACTC-----TTTTC 981

QY 303 SerLeuIleGluLysAlaGlnAspValAspGlnGluLys 316
Db 982 GATTAGTGGAAAGCCGAC-----GATTAGAAAAA 1014
```

```
RESULT 13
US-09-134-001C-1980
; Sequence 1980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyon Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1980
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1980
```

```
Alignment Scores:
Pred. No.: 4,2e-39 Length: 1248
Score: 426.50 Matches: 103
Percent Similarity: 54.57% Conservative: 76
Best Local Similarity: 31.40% Mismatches: 122
Query Match: 18.76% Indels: 27
Db: 4 Gaps: 8
```

```
US-09-943-108a-2 (1-455) x US-09-134-001C-1980 (1-1248)

QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27
Db 298 CAAACCTCCCAAGAACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 357

QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaAspValAsnPheLys 47
Db 358 TTCTTCGAGGCTCTGGAAGAATG-----CTTAATTACTGCGGACGTGTGTTTAAT 408

QY 48 ValValLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 409 ACCGTTATGAATTAACCTGATGAGTACGACAGCAAGCAAGCAAGCAAGCAAGCAAG 468

QY 68 SerLeuThrProGlyGlnGlnValIle---LysIleVal-----GlnAsp 81
Db 469 ACAGAGACTTAAGAGAGTATTAGTTGAGAGATTTAGAAATCTATCATCAAGAGSAC 528

QY 82 GluLeuThrLysLeuMetGlyGlyLysThrSerIleAsnMetSerAsnLysProPro 101
Db 529 GATTATCTCGAGCAATGAATATTGAAGATGGACGTTAAAT-----570

QY 102 ThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeu 121
Db 571 ---GTCATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 627

QY 122 AlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyr 141
Db 628 GCTTATCGTTTTCACAGAA---GGTAAAAAGTAATGTTAGCCGCTGGTGATCACTTC 684

QY 142 ArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr 161
Db 685 AGAGCTGGAGCAATTCACAAATTAACGCTCGGGAGAGACGTTGTGTGTGTGTGTGTGT 744

QY 162 SerGluLysAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLys 181
Db 745 AGTCAAAACGAAAGTCTGACCCCTGCACAGTAGTAGTATGATGCGATTAAATGCGCAAAA 804

QY 182 GluGluHisLeuPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAla 201
Db 805 AATAAGGCGTAGATATTTTATTTGTGTACTGTCAGGACGCTTGCAAAATAAATCTAAT 864

QY 202 LeuMetAsnGluLysGluValLysGluIleAlaLys-----Pro 215
Db 865 TTAATGCAAGAGCTAGATAAAATGAACGTTGTGATTAAATCGTGCAATACCTGATGCCCC 924

QY 216 AsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlu 235
Db 925 CATGAAGCTTTATTATGCTTGGATGCAACACGTTGTCAAAATGCACTTTCACAGCACGT 984

QY 236 SerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThr 255
Db 985 TCATTTAAGGAAGTTACAAATGCTCAGGTATAGTTTAACTAATATTAGACGGTACTGCT 1044

QY 256 ArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGly 275
Db 1045 AAAAGGGGTATTGTTATTAGCAATTCGAAATGAGTTACACATTCACAGTTAAATATGTT 1104

QY 276 MetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIle 295
Db 1105 TTAGCGCAAAAAATGGATGACTTACACCGTTTAAATCCTGAA-----AGCTATGTA 1155

QY 296 LeuGlyMet---GlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314
Db 1156 TATGGATTATTTGCTGATATGATAGACAAATAGATATTTCTCTGAAGAAATCTCTAGA 1215

QY 315 GluLysAlaLysAspLeuGluLys 322
Db 1216 AATTCAATCGTTGTAATCTGGAAGAA 1239

RESULT 14
US-08-858-207A-1980
```

```

; Sequence 180, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Bodysen, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmel, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 180:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-858-207A-180

```

```

Alignment Scores:
Pred. No.: 3,34e-39 Length: 836
Score: 425.00 Matches: 83
Percent Similarity: 76.35% Conservative: 30
Best Local Similarity: 56.08% Mismatches: 35
Query Match: 18.69% Indels: 0
DB: 4 Gaps: 0

```

```
US-09-943-108a-2 (1-455) x US-08-858-207A-180 (1-836)
```

```

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DQ 393 ATGGCATTGAACTTTAAACAGACGCTTTCAGAACGCTTTTAAATAATCTACGTAAAAA 452
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DQ 453 GGAATAAATCTCGAATCTGATGTCACAGAGGCAACCAAGAAATTCGCTTGCCCTGCTC 512
QY 41 GluAlaAspValAsnPheLysValValLysGluPheLleLysThrValSerGluArgAla 60
DQ 513 GAGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DQ 573 GTCGGGATGAGGTCATGTATACACTTAATCTCGGCAACAGATATTATAATCTGTTGAT 632
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DQ 819 ATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1100

```

```

Db 633 GAGGAACACACAGCCGTTTGTAGGTTCTGTATACGGCAGAAATATCATAGTCACCTAAGATT 692
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DQ 693 CCAACATCATCATGATGGTGGTTTACAGGGGCTGTAAACAAACCTTGTCTGTAA 752
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
DQ 753 TTGGCCACAAACCTCAGAAAGAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 812
QY 141 TyrArgProAlaAlaIleAsnGln 148
DQ 813 TATGTCACCTGCCATTCAGACAG 836
RESULT 15
US-09-007-476-1
; Sequence 1, Application US/09007476
; Patent No. 6159949
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6159949el Ftsy
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,476
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GML0079
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1251 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-007-476-1

```

```

Alignment Scores:
Pred. No.: 1.78e-38 Length: 1251
Score: 421.00 Matches: 101
Percent Similarity: 55.59% Conservative: 78
Best Local Similarity: 31.37% Mismatches: 121
Query Match: 18.51% Indels: 22
DB: 3 Gaps: 7

```

```
US-09-943-108a-2 (1-455) x US-09-007-476-1 (1-1251)
```

```

QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27
DQ 289 CAAATTTCCAGACAAATTAATAATTTGATAGCGGATATCGTAAAGTAGATGAAGAC 348
QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaAspValAsnPheLys 47
DQ 289 CAAATTTCCAGACAAATTAATAATTTGATAGCGGATATCGTAAAGTAGATGAAGAC 348

```

```
Db 349 TTTTTCAGACCTTACAGAAATG-----TTAATCAGTCGAGAGTCGGTTTAAAT 399
QY 48 ValValLysGluPheLeIleYsThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 400 ACAGTG-----ATGACGTTAACTGAAGAATTACGTATGGAAGCACAACGACGT 447
QY 68 SerLeuThrProGlyGlnGlnValIleYsIleValGlnAspGluLeuThrLysLeuMet 87
Db 448 AATATTCAAGATACTAGAGATTGCGTGAAGTCATTGTTGAAAAAATCGTAGAGATTAC 507
QY 88 GlyGlyGlu-----AsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
Db 508 CATCAAGAAGATGATAAATTCAGAGCTATGACATAGAACAGATGCTCGTTAAATGTCATT 567
QY 105 MetMetValGlyLeuGlnGlyAlaGlySyrThrThrAlaGlyLysLeuAlaLeuLeu 124
Db 568 TTAATGTCGGTGTGAATGCTGGTAAACACACAACTTGGAAATTAGCT--TAC 624
QY 125 MetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAla 144
Db 625 CCATATAAATGAAGGTAAAAAGATATGTTAGCTCGGGCGATACCTTTAGAGCGGGT 684
QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGly 164
Db 685 GCTATTGATCAATTGAAGCTTTGGGGCGACGCTGTGGTGTAGATGTAATTAGCCAAAGT 744
QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
Db 745 GAAGGTCTGATCCACTGCTGTTATGTATGATGCGGATTAATGCCGCTAAAAACAAGGT 804
QY 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
Db 805 GTTGATATTTAATCTGTGATACCGCTGGACGTTTACAAAATAAAACAATCTAATGCCAA 864
QY 205 GluLeuLysGluValLysGluIleAlaLys-----ProAsnGluIle 218
Db 865 GAATTAGAAAAGCTTAAGCGTGTAAATTAATCGACAGTCGCCAGATCGCCCTCATGAAGCA 924
QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 925 TTACTATGTTAGATCTACACTGCTCAGATCGGTGATGCTGCTCAGACCTAGAACTTTAAA 984
QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 985 GAAGTACAAATGTTACAGTATTGTTAATACGAATTAGATGGTACAGCCAAAGGTGGT 1044
QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleIysPheValGlyMetSerGlu 278
Db 1045 ATCGTATTAGCCATTGTAATGAATGCACATCCAGTTAAATATGATAGTTTAGGTGAG 1104
QY 279 LysLeuAspGlyLeuGluLeuPheHisProGluArg-----MetAlaSerArg 294
Db 1105 CAATTAGATGACTTACACCACTTAATTAACCTGAAGTTATGCTACGGCTTATTCGCTGAT 1164
QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGln 314
Db 1165 ATGATTGAACAAAATGACGGAATAACACACAGTGAATAAT---GATCAAAATTGTACACAAA 1221
QY 315 GluLys 316
Db 1222 GAARAG 1227
```

Search completed: February 25, 2003, 03:14:46
Job time : 109 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:59:07 ; Search time 89 Seconds
(without alignments)
2871.303 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274
Sequence: 1 MAFEGSERLQATMQRGK.....CKGKRNQNMQLKGMNLPF 455

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Egapop 6.0 , Egapext 7.0
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09943108/runat_21022003_151633_1953/app_query.fasta_1.647
-DB=PublishedApplications.NA -QFMT=fastap -SUFFIX=rnbp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09943108.ecgnl_1_80_@runat_21022003_151633_1953
-NCPU=6 -ICPU=3 -NO_XLPUY -NO_MMAB -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications.NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	2261	99.4	1368	10	US-09-815-242-8968	Sequence 8968, Ap
2	2235	98.3	1365	10	US-09-815-242-4365	Sequence 4365, Ap
3	2235	98.3	1368	10	US-09-815-242-8030	Sequence 8030, Ap
4	1418.5	62.4	1416	10	US-09-815-242-3890	Sequence 3890, Ap

Sequence 6764, Ap
Sequence 37, Appli
Sequence 1847, Ap
Sequence 9156, Ap
Sequence 9427, Ap
Sequence 1, Appli
Sequence 2255, Ap
Sequence 6166, Ap
Sequence 1945, Ap
Sequence 7850, Ap
Sequence 6868, Ap
Sequence 9991, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 14721, A

ALIGNMENTS

RESULT 1
US-09-815-242-8968
; Sequence 8968, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8968
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8968

Alignment Scores:
Pred. No.: 5,28e-245 Length: 1368
Score: 2261.00 Matches: 452
Percent Similarity: 99.78% Conservativeness: 2
Best Local Similarity: 99.34% Mismatches: 1
Query Match: 99.43% Indels: 0
DB: 10 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-815-242-8968 (1-1368)

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTGAAGGGTTATCAGACGGCTTGCAAGCGACGATGCAAAAAATGCGTGAAG 60

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACCTTACTGAACCTGATTAAGATAATGATGGTGAAGTAGATGAGCTTACTT 120

QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAACTTTAAAGTGTAAAGCAATTTATTAACACAGTATCAGACGGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTCCGATGATGCAATCATTAACACAGGCAACAGTTATTAATAATAGTTCAA 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGATTAACCGAGTGTGATGGGTGAGAAATACGTCGATTAATATGTCAAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTGTATGATGTTGTTTACAAAGTCTGTGTAACACACAACTGCAGGTAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATGCGTAAATAATACAAACAAACAACTATGTTAGTTCAGCAGATAT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGTCACAGCAGGATAAATCAATTAACACAGTACGAGCAAAATGATATCCTGTA 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGTAATAATGATTAATCAATGCT 540

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AARGAAGAACATTTAGACTTTGTATATCAATGATACAGCAGGCTGATTAACATCATGAA 600

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGACGATTAATAAGAGATTAAGACATTCCTTAACCAACAGAAATATGTTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTCGATTCATGAGGGTCAAGATGCTGCTCAATGTTGCAAGATCTTTGACGATCAA 720
```

```
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTTACCTTAACATTAATAGATGGTGATACACAGTGGTGGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCTGTTGCGTGACACAAACCAATTAATTTGTTGGTATGAGTGAAGATTTA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTTACAGCTATTCCATCTCTGACGGTATGTCATCAGTATTTAGGTATGGGTGAT 900

QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAAATGAAAAAGCCACACAGATGTGATCAAGAAAAAGCAAAAGATTTA 960

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluLeuAspGln 340
Db 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTATAGAACAACTTGTATCAG 1020

QY 341 VallLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCCTGATGATATTAATAAATGATTCAGGTATGAATAAATG 1080

QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGGCTAGATTAAGCTTAATATGATGTAAGAAAAAATCCACACACATTAATAAAGCATATC 1140

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGGCGCTGAAAGAAACAAATCCACACACATTAATAATGATCAAGTAAAG 1200

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTAAAGGGTCTGCTGCTTCATTAACAGAAAGTCAATCGTTGATGAACAATTT 1260

QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
Db 1261 AACGATATGAAGAAAAATGATGAACAGTTCAGTGGCGGTAAGGTAAGGTAAGGTAAG 1320

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CCAATCAATGACAAATTAATGTTAAAGGTATGAATTTACCGGTTT 1365

RESULT 2
US-09-815-242-4365
; Sequence 4365, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4365
; LENGTH: 1365
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4365

Alignment Scores:
Pred. No.: 4,42e-242 Length: 1365
Score: 2235.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 10 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-815-242-4365 (1-1365)

QY 1 MetAlaPheGluGluSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGGCAATTTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATGCGTGTAAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGTAACCTACTGAGCGTGATATAAGATATATGATCGGTGAAGTAAGATTAGCGTTACTT 120
QY 41 GluAlaaspValasnPheLysValLysValLysGluPheIleLysThrValSerGluArgAla 60
DB 121 GAGCGTCAGCTAAACTTTAAAGTGGTAAAGAAATTTATTAAACACAGTATCAGAACGCGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
DB 181 TTAGGTTCGATGTAATGCAATCATTAACACCGAGGCAACAAGTTATTAAATAGTCAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
DB 241 GATGAATTAACGAGCTGATGGGTGAGAAATATATATGATTAATGTCATTAACCA 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CCTACTGTGTTATGATGGTGGTTTACAAGTCTGCTGTAAACAACTGCAGGTAAA 360
QY 121 LeuAlaLeuMetArgLysIleLysIleLysIleLysIleLysIleLysIleLysIleLysIle 140
DB 361 TTAGCATTTATGCGTAAAGAAATACACAAACAAACCTATGTTAGTTCGACGAGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
DB 421 TATCGTCCAGCAGGATAAATCAATTAACAACAGTAGGGAACAATTTGATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
DB 481 TACAGTGAAGGAGATCAAGTAAAGCCCAACAATTTGTAATGCAATTAACAACTGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
DB 541 AAGAAGAATTTAGACTTTGATCATTTGATACAGCAGGTCGATTAACATTCGATCGATGA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
DB 601 GCATTTGATGAATGAATTAAGAAGTAAAGACATTTGCTAAACCAACAAAGAAATTATGTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
DB 661 GTTCTGATTCATGACGGGTCAAGATCGTCAATGTTGCAGAACTCTTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
DB 721 CTTGATGTCACAGGTGTACCTTAACATAAATAGATGGTATACACGCTGGTGGTGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
DB 781 TTATCATTTCCGTTCCGTGCACAAAAACCAATTAATTTGTTGTTATGATGAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
DB 841 GATGGTTTAGAGCTATTCCATCCTGAAACGATATGGCATCACGATTTTCAGGCGCTGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
DB 901 GTCTCAAGGTTAATTGAAAAAGCCACACAGATGTGGATCAAGAAAAAGCAAGAAITTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
DB 961 GAGAAAAAGATCGCGAGTCATCAITTTAGATGATTTTGTAGAACACTTGTCTAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
DB 1021 GTGAAAAATTTAGGACCACCTGGATGATATTATTAAATGATTCACGATATCAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
DB 1081 AAAGGCTAGATAAGCTTAATAGTGAAGAAACAAATTTGATCATATTAAAGCGATTATC 1140
QY 381 GluSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
DB 1141 CAGTCAATGACCGCGCTGAAAGAAACAATCCAGACACATTTGATGATCACTGATTAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
DB 1201 CGTATGCTAAAGGTTCTGGTCTTCATTACAAGAAAGTCAATCGTTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
DB 1261 AACATATGAAGAAATGATGAACAATTCACCTGGCGGTAAGGTAAGGTAAGGTAAG 1320
QY 441 ArgAsnGlnMetGlnAsnMetLysLysGlyMetAsnLeuProPhe 455
DB 1321 CGCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

RESULT 3

US-09-815-242-8030
; Sequence 8030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8030
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8030

```

```

Alignment Scores:
Pred. No.: 4,44e-242      Length: 1368
Score: 2235.00           Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28%      Indels: 0
DB: 10                    Gaps: 0

```

US-09-943-108A-2 (1-455) x US-09-815-242-8030 (1-1368)

```

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTTTGAAGCTTATCAGAAAGCCTGCAAGCGACGATGCAAAAAATCGTGAAG 60
Qy 21 GlyLysLeuThrGluAlaAspLysLysLysMetMetArgGlnValArgLeuAlaLeuPhe 40
Db 61 GGTAACTTACTGAGCTGATATAGATATATGATGCTGAAGTAGATTACGCTTACTT 120
Qy 41 GluAlaAspValAsnPhelysValLysGluPheLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAACTTTAAAGTGTGTAAGAAATTTATTAACACAGTATCAGAGCGCA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80
Db 181 TTAGTTCGATGTAATGCAATCATTAACACGAGCGCAAGATTTATTAATAGTTCAA 240
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGATTAACGAAGTTGATGGGTGGAGAAATATATCGATTATATGTCAAATAACCA 300
Qy 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGGTGTGTTTACAAAGTGTGTGTAACAACTCCAGGTAAA 360
Qy 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATCGGTGTAATAATACACAAAAACCTATGTTAGTTCACACAGATAT 420
Qy 141 TyrArgProAlaAlaLysGlnLeuGlnThrValGlyLysGlnLysLysLysProVal 160
Db 421 TATGCTCCAGCAGCGATAATCAATTACAAACAGTAGGAAACAAATTTGATTTCTGTA 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
Db 481 TACGTGTAGAGAGATCAGTATAGCCACAAATTTGTAATGATGATTAACAACTGCT 540
Qy 181 LysGlnGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisLysAspGlu 200
Db 541 AAAGAGAACAATTTAGACTTTGTAATCATTTGATACAGCAGCTCGATTACACATCATGAA 600
Qy 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysLysProAsnGluLeuLeu 220
Db 601 GCATTTGATGATGATTAATAAGAAAGTAAAGACATTTGTAACCAACGAAATTTATGTA 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATTCATGACGGTCAAGATGCTGTCATGTTGAGAACTTTTGACCATCAA 720
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyLysAla 260
Db 721 CTTGATGTCACAGGTGTTTACCTTAACATAAATAGATGCTGATACAGCTGTGTGACAGT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280

```

```

Db 781 TTATCTATTGTTCCGTTGACACAAAAACCAATTAATTTGTTGGTATGAGTAAAAAGTTA 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGTTTATAGAGCTATTCATCCTGAACGTATGTCATCAGCTATTTTCAGGCTGGTGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTCAAGGTTTATTTGAAAAACGCAACAGATGTGGATCAAGAAAAAGCAAGATTTA 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
Db 961 CAGAAAAAGATGCGGAGTCAATCTTACITTAGAAGATTTTATAGACACATTTGATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGGACCACTGGATGATATTATTAATAATGATCCAGGTATGAATAAATG 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisLysLysLysLysLys 380
Db 1081 AAAGGTCTAGTAAGCTTAATATGATGAGTGAAGCAAAATGATCATATTAAAGCGATATC 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGCGGCTGAAAGAAACAATCCAGACACATTTGATGATCAGTAAAAAG 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTCTTAAGGTTCTGCTGTTCAATTAAGAAAGTCAATCGTTTATGATGAACAAAT 1260
Qy 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLysLys 440
Db 1261 RACATATGAGAAATGATGAACAAATTTGATGCGGTAAAGGTAAAGGTAAAGGTAAA 1320
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365

```

RESULT 4

```

US-09-815-242-3890
; Sequence 3890, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 3890
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3890

Alignment Scores:
Pred. No.:      3,17e-150      Length:      1416
Score:          1418.50        Matches:      273
Percent Similarity: 77.54%     Conservative:  86
Best Local Similarity: 58.96%   Mismatches:   89
Query Match:     62.38%       Indels:       15
DB:              10           Gaps:         2

US-09-943-108a-2 (1-455) x US-09-815-242-3890 (1-1416)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 1 ATGGCTTTTGAGAGTTTACAAACCGCCTACACAGGCAATGAGTAAATCCGTCGTAAG 60
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 21 GlyLysLeuThrCluAlaAspIleLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 61 GGAAGAGTTCCGAGCGCGAGCTAAAAGAAATGATCGAGAAATCCGTTGGCTTATTA 120
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 121 GAAGCGAGCTTAATTACAGTGGTCAAGAGATTTCACAAACCGGTCAGAGAACGGGCA 180
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 61 LeuGlySerAspValMetLysSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 181 GTAGGAGTCGAAGTATTAGAAGCTTATCACCAGCCCAACAATTTGTAATAATTTGTTGAT 240
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 81 AspGluLeuThrLysLeuMetGlyGlyLysThrSerIleAsnMetSerAsnLysPro 100
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 241 GAAGATATACGAAATCTTAGGTCAGAAACCGGTTCAACTGAATAAATTCCTCAAAAATC 300
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 301 CCGACAGTGATATGATGACAGGGTTACAGGGGCTGGTAAACACACTTTTACTGGTAA 360
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 361 TTAGCAAAACACTTAATGAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 141 TyrArgProAlaAlaAlaLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 421 TATCGTCCAGCAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 481 TTTCATATGGACACAGCTTAATCCAGTGGAAATTTGCTCAGGCTTAGCATTAGCA 540
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 541 AAAGAAAGAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 601 GCITTAATGACGATTAATGAAACAAATTAAGAGTTTGGCTTAATCCCAATGAAATTCGTGA 660
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 221 ValValAspSerMetThrGlyGlnAlaValAlaAsnValAlaGluSerPheAspGln 240
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 661 GTTGTGATGCGATGACGGGCGAGATGCTGCTCAAGTTGACAGATAGTTTATGAAACAG 720
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 721 CTTGCAATTAATGCGGGTGTATTATACCAAAATGACCGCGATACGCTGGGGGGCGTGGC 780
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 781 CTGTCAATTCGGGCGAGTAAAGCGCTCCGATTAATTTCTCGGTTCTGGTGAATAATTA 840
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
```

; SEQ ID NO 6764
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1434)
US-09-815-242-6764

Alignment Scores:

Pred. No.: 3,23e-150 Length: 1434
Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-6764 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgGluGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCTTTGAGAGTTTAAACAACCCGCTACCAACAGGCAATGAGTAAATCCGTCGTAAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAGAGTTTCGAAAGCGGACGTAAAGAAATGATGCGAGAAATCCGTTGGCTTTATTA 120
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAAGCCGAGGTTTAAATTCAGAGGGTCAAGAAATTTCAAAAACGCTCAGAGAGCGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTAGAGTCGAGATATTAGAAAGCTTATCACCGCCCAACAATTTGTAATTTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAAGAAATTAACGAAACGTTAGGTTCAGAAACGGTTGAACGTAATAATCCCAAAATC 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACAGTGATGATGATGACAGGGTTACAGGGGCTGTTGTAACAACTTTTACTGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCAAAACACTTAATCAAACTGAAACGCTGCTCCGCTTTTAATCGCTGGTGAGGT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATTGATGATGAGGTGTTAGGTGCAACAAATAGAAAGTTCCCGGT 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTGATATGGAAACAGATGCTAATCCAGTGGAATTTGCTCAAGGTTTAGCAATTAGCA 540
QY 191 LysGluGluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAAGAAATGATTATGTTCTTAATGATGACGGCGCGCTTTTACACATTGACGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCTTAAATGAGCAATTAACAAATTAAGAGTTGGCTTAATCCCAATGAAATTCGTGA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATCGATGACGGGCGAGATGCTGTCACGTTGACAGATAGTTTAAATGAACAG 720
QY 241 LeuAspValThrClyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGGAATTAAGGGTGTGTTATTACCAAAATGACGGCGGATCTCGTGGGGGGGCTGCG 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

Db 781 CTGTCAATTCGGCAGTAAACGGGCGCTCCGATTAAATTTGTCGGTTCTGGTGAATAATTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACCGATTTAGAAATTTTCCATCCGATCGTATGCGAGTCGTATAGGTATGGGGAC 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 AGTTGACGCTAATTGAAAACCGCAACAGATTACAGTACAGAGAAAAGCAGAGAACTT 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GGTCAAAAATGAAGAAAACAGTTTTTACGTTAAACGATTTCATTGACCAATTTGATCAA 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTATGGCAGTGGGACCGGATTGAAGACTTATTAATAATGATCCCTGGATGAGTAACATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCTGTATTTGAAATGTCAAAAGTCGATCCAAAAGATGTGGCAGGAAACGGCGATGGTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CTTATCAATGACCCCTGCAGAACGTTGAAATCTGATCTATTAAATCCTAGTCGCCGTCGC 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 AGAATTTGACGTCGTTTCAGGAAATAGTGTGTTGAAGTCAATCGTATGATTAAACAATT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThr----- 431
Db 1261 AAAGAAATCAAAAATAATGCAACAAATGTCCAAAGGGGATATGAACATTCCTGGTATG 1320
QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMetGlnAsn 446
Db 1321 GATCAAAATGCTAGTGGCGCGGTTTAAAGCAAGTATAGTTAAAATG---GCCATGATCGT 1377
QY 447 MetLeuLys 449
Db 1378 ATGATGAAG 1386
RESULT 6
US-09-070-927A-37
; Sequence 37, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06

```

; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-070-927A-37

```

Alignment Scores:

```

Pred. No.:      3,83e-149      Length:      6729
Score:          1417.50        Matches:      273
Percent Similarity: 77.54%      Conservative: 86
Best Local Similarity: 58.96%    Mismatches: 89
Query Match:      62.34%       Indels:      15
DB:               10          Gaps:         2

```

US-09-943-108a-2 (1-455) x US-09-070-927A-37 (1-6729)

```

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3722 ATGGCTTTTGAGAGTTTAACAAACCGCTACAGCGCAATGAGTAAATCCGCTGTAAG 3781

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3782 GGAAGAAGTTCCGAAGCGCAGTAAAGAAATGATCCGAGAAATCCGTTTGGCTTTATTA 3841

QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3842 GRAGCGGAGTTTATTCACAGTGGTCAAGATTTTCAAAACGCTGTCAGAGACGGGCA 3901

QY 61 LengSerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3902 GTAGGAGTCGAAGTATTAGAAGCTTATCACCAGCCCAACAAATGTGTAATAATTTGTGAT 3961

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3962 GAGAATTAACGAAACCTTAGGTTCAGAACGGTTGAACCTGAATAAATCTCCAAATATC 4021

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4022 CCGACAGTGATTATGATGACAGGGTTACAAAGGGGCTGTGTAACAACTTTTACTGGTAA 4081

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4082 TTAGCAAAACACTTAATGAAACCTGAAACCGCTCGCTGCTTTTAAATCGTGTGACGTT 4141

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4142 TATCGTCCAGCAGCGATTGATGCTAGTGAAGGTTTATAGTCAACAATTAGAAGTCCCGTT 4201

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4202 TTTGATATCGGACAGATGCTAATCCAGTGGAAATGTTCCTCAGGGTTAGCATTAGCA 4261

QY 181 LysGluLuhHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4262 AAAGAAAAGAAATGATTATGCTTAATGTATACGGCGCGCTTTACACATGACGAA 4321

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4322 GCITTAATGGAGAAATGAACAAATTAAGAGATTGGCTGATCCCAATGAATTCGTGTA 4381

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240

```

```

Db 4382 GTGTTGTGATGCGATCAGCGGSCAAGATGCTGTCAACGTTGCAGATCTTTAATGACAG 4441

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4442 CTTGGAATTAATGCTGGGTTGTTATTACAAAATGGATGGCGATCTCGTGGGGGGGTGG 4501

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4502 CTGTCAATTCGGGCGAGTAACGGGTGCTCCGATTAATTTGCGTTCTGTGAAAAATTA 4561

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4562 ACCGATTAGAAAATTTCCATCCCGATCGTATGTCGATCGTATCCTAGTATGGGGGAC 4621

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4622 ATGTTGACGCTAATTTGAAAAAGCGCAACAGATTACGATGAGAAAAAGCAGAAAGACTT 4681

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4682 GCTCAAAAATGAAGAAACACGTTTGTGACTTTAACGATTTTCATTGAGCAATTTGGATCAA 4741

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4742 GTTATGGCGATGGGACCGATTGAGACTTATTAATAATGATCCCTGGAAATGACATACAG 4801

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4802 CTGTGATTGAAAATGCAAAATCGATCCAAAGATGTGGCAGCAAAACGGCGGATGTC 4861

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4862 TTATCAATGACCCCTGCAGACGCTGAATAATCCTGATCTATTAAATCTAGTCGCCCGGC 4921

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4922 AGAATTGACGCTGTTCTCAGGAATAGTGTGGTGAATCAATCGTATGATTAACAAATTT 4981

QY 421 AsnAspMetLysLysMetMetLysGlnPheThr----- 431
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4982 AAAGAAATCCAAAAAATGATGCAACAATTCGCAAAAGGGGATATGAACATTCCTGCTATG 5041

QY 432 -----GlyGlyGlyLysLysLysGlyLysArgAsnGlnMetGlnAsn 446
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 5042 GATCAAAATCTTAGTGGCGCGGTTAAAGGCAAGCTAGGTAAATG---GCCATGAATCGT 5098

QY 447 MetLeuLys 449
    |||||  |||||
Db 5099 ATGATGAAG 5107

```

RESULT 7

```

US-09-974-300-1847
; Sequence 1847, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1847
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Bacillus licheniformis

```

US-09-974-300-1847

Alignment Scores:

Pred. No.:	1,566-148	Length:	1136
Score:	1402.00	Matches:	270
Percent Similarity:	87.30	Conservative:	60
Best Local Similarity:	71.43	Mismatches:	48
Query Match:	61.65	Indels:	0
DB:	10	Gaps:	0

US-09-943-108A-2 (1-455) x US-09-974-300-1847 (1-1136)

QY	1	MetalApheGluLysSerGluArgLeuGlnAlaThrMetGlnLysMetArgLys	20
Db	1	ATGGCATTCGAAGGATTACCGCCGACCTGCACGAACGATCTCAAAAATCCGCCGAA	60
QY	21	GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
Db	61	GGAAAGGTACGGCAGACGTCNAAGAGATGATGCCGGAATCCCTTTAGCCCTCTT	120
QY	41	GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla	60
Db	121	GAAGCGGACGTCAATTTTAAGTCGTCAAAGATTGTTGTAAAAAGGTAAAGCAACGAGCT	180
QY	61	LeuGlySerAspValMetLeuSerLeuThrProGlyGlnGlnValIleValGln	80
Db	181	GTCCGACAACACGCTTATGAAGAAGCTTGACGCCCGCACAGCAGGTCAATAAAGTGT	240
QY	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
Db	241	GAAGAGCTGACGAGCTGATGGCGGGAGGAACGAAGATCCCGCTGCACAAAGGTCG	300
QY	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGLys	120
Db	301	CGACCGCTCATCATGATGSGCTGCTGCAAGGGCGGTAAACGACGTCGACCGGAAG	360
QY	121	LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	361	CWTGCCAACCTGCTCGCAAAAACATACCOCGAATCCGCTGCTGGCAGCAGACATC	420
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	421	TACCGCCCGCGCGCATCAACACCTGGAAAGCGCTCGCAAGCAGCTCGATATCCCG	480
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	481	TTTTACTGGTGACAAAGTCAGTCCGGTGGAGATTGCCAAGCAGGCCATCGAAAGGCG	540
QY	181	LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	541	AAAGAAGATCATCACGACTACGTTCTCATCGATCGCGGGAGCCCTTCATATGATGA	600
QY	201	AlaLeuMetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGluIleMetLeu	220
Db	601	GAGCTGTAGTATGAGCTTGACAGGTAAGAAACCCGCTCAGCCGGAAGAGATTTCC	660
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValIaGluSerPheAspAspGln	240
Db	661	GTTCGTGTACTCCATGACGGGTCAAGACCGGTCAATGTGCCAAAGACCTTAAACGAC	720
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla	260
Db	721	CTCGGTGTGACGCTGTCATCTTGACGAAGCTTGACGCGATACTAGAGCGGAGCGGCC	780
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	781	TTGTCAATCCGCGCCGTGACAACACCGCATCAATTCGCCGGAATGGCGCAAAAGCTT	840
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleuGlyMetGlyAsp	300
Db	841	GATGCGCTGGAGCCGTTTCATCTCTGAACGAGTAGGCTTCGCGGATATCCGAATGGCG	900
QY	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu	320

```

Db 901 GFGCTCACCTTATGCGAAGAGCCGAGGCCAACGTAGACGAAGAAAGGCGAAGAACTC 960
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Qy 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Db 961 GAGCAAAAATGAAAATAATGACGCTTACGCTCGACGACCTTTTGGACACGCTCGGACAG 1020
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Db 1021 GTCGCAATATGGCGCGCTTGAAGACCTGATTCAAATGATGCCGGCGCAGCAAAATG 1080
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAla 378
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Db 1081 AAAGGCTTAAAAACGCTGAAGGTGTAGTAAAAACAGCTCAGCCATATCGAAGCG 1134
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||

RESULT 8
US-09-815-242-9156
; Sequence 9156, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191.078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206.848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207.727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242.578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253.625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257.931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269.308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9156
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1572)
US-09-815-242-9156

Alignment Scores:
Pred. No.: 1.28e-137 Length: 1572
Score: 1307.00 Matches: 253
Percent Similarity: 74.73% Conservative: 90
Best local Similarity: 55.12% Mismatches: 102
Query Match: 57.48% Indels: 14
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-9156 (1-1572)

Qy 1 MetaLpHeGluGlySerGlnArgLeuGlnAlaThrMetClnLysMetArgGlyLys 20
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Db 1 ATGGCAATTTGAAAGTTTAAACGAACGTTTGCAGAACGCTCTTTAAATAATCTACGTTAAAAA 60
      |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:|||||
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

```

Db 61 GGAARAATCTGATCTGATGTCACAGAGCGCAACCAAGAAATTCGCTGGCCGTGC 120
Qy 41 GLUALASPVALASPHELYSVALVALYSLUPLHEILYSLYTHRSERGLUARGALA 60
Db 121 GAGCGGAGGTGGCTTCCTGCTGTTGTAAGAGACTTTATCAAGAAAGTTCGTGAGCGTGA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTGGGATGAGGTCAATGATACACTTAATCTCGCGCAACAGATATATAAATCGTTGAT 240
Qy 81 AspGlnLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAGAACTGACAGCGCTTTAGGTCTGATACGGCAGAAATTTATCAACTCACCTAGATT 300
Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACCATCATGATGTTGGTTTACAAGGGCTGTGTAAACAACCTTTGCTGTGATAA 360
Qy 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTGGCAACAACACTCAGAAAGAGAAATGCTGCTCTTGTGATGTTGGCGGGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGCTCCAGCTGCCATGACCACTTAAGACCTTGGGACACAGATTGATGTCCTGTC 480
Qy 161 TyrSerGlnGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTGCTACTTGGCAAGAACTACCAAGCTGTTGAGATTGTACGTCAAGTTTGGAGCAAGCC 540
Qy 181 LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 CAACTAAATCAATACGATATGTTCTGATGATGACGCGGTCGTTTGCAGATTGATGAG 600
Qy 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGATGAGTTCGTGATGTGAAGCATTTGGCTCAACCAATGAATCTTGCTT 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTGGTGTGATGCTATGCTTACCAAGATTGATGGCCATPACCTGGTGGTGTGCT 780
Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 TTGGAAGTCACTGGGTGATCCTTACCAAGATTGATGGCCATPACCTGGTGGTGTGCT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTCTGCTGCATACCTACTGGAACCAATCAAGTTCACGTGTACAGGTGAAAGATT 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATTAAGAACCTTCACCCAGACCGCATGCTAGCCGTATCTGTGTATGGGGAT 900
Qy 301 ValLeuSerIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACTTTGATGAGAAAGCTTCACGAAATACGATGAGAAACAAAGCCCTTGAATG 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATGCGGCAAAACACCTTTGATTTAATGATTTTCATCGATCAATTAGATCAG 1020
Qy 341 ValIysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGCCCATGTAAGAGACTTGCTCAAGATGATTCAGGTATGCCCAACAT 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAAAACATGAAGGTGGATGAACGCCAGATTGCTGCTAAACGTGGCATGTG 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400

Db 1141 TCTTCGATGACACCTGAGAGCGTGAAAACCCAGATTTGTTAAATCCAGCCGTGCCGT 1200
Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGGTTCGGAATACATACATCTGCTCAAGTCAATAAATTCATCAAGGACTTT 1260
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysGlyLys 440
Db 1261 AACAGCTAAACACATCATCAGGGGTGTTATCTCTGGG----- 1299
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAACAAATGGGATTAATCCAAATACCTTCCT 1350
RESULT 9
US-09-815-242-9427
; Sequence 9427, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1395)
US-09-815-242-9427
Alignment Scores:
Pred. No.: 1.39e-137 Length: 1395
Score: 1306.00 Matches: 252
Percent Similarity: 74.73% Conservative: 91
Best Local Similarity: 54.90% Mismatches: 102
Query Match: 57.43% Indels: 14
DB: 10 Gaps: 2
US-09-943-108A-2 (1-455) x US-09-815-242-9427 (1-1395)
Qy 1 MetalapheCluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGCGATTTGAAGATTTTACAGAACGTTTGCAGAACGCTTTTAAAAATCTACGTAAAAA 60
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetArgGluValArgLeuAlaLeuPhe 40

```
Db 61 GGAATAATCTCTGAATCTGATGTCCAGAGCCAAACCAAGAAATTCGCTTGCCCTGCTC 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCCGAGCGTGCCTGCTGTGTAAGGACTTTATCAAGAAAGTTCGTGAGCGTGA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysLeuValGln 80
Db 181 GTGCGGATGAGTGCATGATCACTTAATCTCGCAACAGATTTATTAATCTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAGGAACATGACCGCTTTTATGTTCTGATACGGCAGCAAAATATCAAGTAAAGATT 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACCATCATCATGATGTTGTTTACAGGGCTGTGTAACACACCTTTGCTGGTAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
Db 361 TTGGCCACAACTCAAGAAAGAAAGAAATGCTGCTCTTGTGATTCGGCGCGATATT 420
QY 141 TyrArgProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGCTCAGCTGCCATTCACAGCTTAAGACCTTGGGACACAGATTTGATGCTGCTC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTGCACTTGGACAGAGATCACAGCTGTGAGATTTGATCTCAAGTTTGGACAGGCC 540
QY 181 LysGluGlnLysLeuAspPheValIleAspThrAlaGlyArgLeuHisLeuAspGlu 200
Db 541 CAACATATCATACACGATCATCTGATGATGATGATGATGATGATGATGATGATGATG 600
QY 201 AlaLeuMetAsnGlnLysGluValLysGluLeuAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 TTGGAAGTGAATGGGCTCATCTCCACAGATTTGATGATGATGATGATGATGATGATG 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTGTGTCATCACTTACTGGAACCAACCAATCAAGTTCAGTTCAGGTGAAAGATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGACATGAAACCTTCCACCCAGCCGATGCTAGCCGATGCTGATGATGATGATGATG 900
QY 301 ValLeuSerLeuIleGlnLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 321 GlnLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGGAAGATGCGGAAACACCTTGTATTTAATGATTCATCATCAATATGATGATGAT 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAATATGAGGCGGATGGAAGCTTGTCAAGATGATTCAGATGATGATGATGATGAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCACTTCAAAACATGAAGTGGATGAACGCCAGATTCCTCGTAAACGTCGCTATGTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACATGGAAGAACCTGAAACCCAGATTTGTTAAATCCAAAGCCGTCGCGCT 1200
```

```
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGCTGCTGGAATACATTCCTCGAGCTCAATTAATTTATCAAGACTTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACACGGCTAAACAGCTCATCGAGGCTGTATGCTGGG-----1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAAATGATGAAGCAATGGGAGTAAATCCAAATAACCTTCT 1350
RESULT 10
US-09-814-041A-1
; Sequence 1, Application US/09814041A
; Patent No. US20020103104A1
; GENERAL INFORMATION:
; APPLICANT: CHEEVER, CHRISTY
; APPLICANT: FECTEAU, DOUGLAS A.
; APPLICANT: LI, HU
; APPLICANT: PAYNE, DAVID J.
; APPLICANT: STEEL, ANGELA
; APPLICANT: WANG, LEI
; TITLE OF INVENTION: METHODS USING THE SRP POLYNUCLEOTIDES
; AND POLYPEPTIDES COMPOUNDS MODULATING THEIR ACTIVITY
; FILE REFERENCE: GM50069
; CURRENT APPLICATION NUMBER: US/09/814,041A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/191,008
; PRIOR FILING DATE: 2000-03-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-814-041A-1
```

```
Alignment Scores:
Pred. No.: 1.7e-136 Length: 1569
Score: 1297.00 Matches: 251
Percent Similarity: 74.51% Conservative: 91
Best Local Similarity: 54.68% Mismatches: 103
Query Match: 57.04% Indels: 14
DB: 10 Gaps: 2
```

US-09-943-108A-2 (1-455) x US-09-814-041A-1 (1-1569)

```
QY 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTGAAAGTTTAAACAGACGCTTTCGAGAACGCTTTTAAATATCTAGCTAATA 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAATAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGGCGGACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTGCGGATGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 81 AspGluLeuThrLysLysMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAGGAACATGACCGCTTTTATGTTCTGATACGGCAGCAAAATATCAAGTAAAGATT 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCACCATCATCATGATGTTGTTTACAGGGCTGTGTAACACACCTTTGCTGGTAA 360
```

```
QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTGGCCAAACAACACTCAAGAAAGAGAAATGCTGCTCTTATGATGGCGGATAT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCTGCCATTGACCAGCTTAAGACCTTGGGACACAGATGATGTCGCTGC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTGGCACTTGGACAGAGATACCAAGCTGTTGAGATGTACGTCAAGTGTGGACGAGCC 540
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 CAACACTAATCAATACGATGATCTTGATGATACCTCGGGGTCGTTGCAGATGATGAG 600
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220
Db 601 CTCCTCATGAATGAGCTTCGTGATGTGAAAGTATTGGCTCAACCAATGAATCTTGCTT 660
QY 221 ValValAspSerMetThrGlyClnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTCGTTGATGCTATGATGCTGCAGAGACAGCAATGTTGGCGCTGATGTTAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
Db 721 TTGGAGTGCATGGGGTCACTTACCAAGATGATGCGATGATGCGGTGATGCTGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTCTGTCGTCACATCACTGGAACCAACCAATCAAGTCACTGTTACAGGTGAAAAATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACAGATATCGAAACCTTCCACCCAGACCGTATGCTTAGCCGTATCCCTTGGCATGGGGAT 900
QY 301 ValLeuSerIleLeuLysAlaGlnLysAspValAspGlnLysAlaLysAspLeu 320
Db 901 ATGCTCATTGATGAGAAAGCTTCACAGGAATACGATGACAAAGAACCCCTTGAATG 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATGCGGAAACACCTTTGATTTAATGATTTCATCGATCAATTAGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTCGAAATATGGGCGCATGGAAGACTTGCTCAAGATGATCCAGGTATGGCCACAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCACTTCAAAACATGAAGTGGATGAACGCCAGATGCTGTAAACGTGCCATTGTG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATCAGATCTCGAAGACGTGAAACCCAGATTTGTAATCCAAAGCGCTCGCGT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTGCTGGGTTCTGGAATATACATTCGTGCAAGTCAATAAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLys 440
Db 1261 AACCAGCTTAAACAGCTCATCGAGGCTGTTATGCTCGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATATGATGACGAAATGGGATTAATCCAAATACCTTCT 1350
RESULT 11
US-09-738-626-2255
; Sequence 2255, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
```

```
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2255
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2255
```

```
Alignment Scores:
Pred. No.: 5,84e-122 Length: 1641
Score: 1168.00 Matches: 237
Percent Similarity: 66.67% Conservative: 85
Best Local Similarity: 49.07% Mismatches: 125
Query Match: 51.36% Indels: 36
DB: 5 Gaps: 5

US-09-943-108a-2 (1-455) x US-09-738-626-2255 (1-1641)
```

```
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGAGTCACTGCTCGGATCGGTTGAATAGCGCGCTTTCGGCGCTCGCGGCAAGGAAAG 63
QY 23 LeuThrClnAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTCACGAGGAGAGACATCAATGCACACGAGATCCGCTCGCGCTGCTGCGGAGGT 123
QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GAGCTTCATTAAACGGTGTTCGTGCTTCATTAAACGGAATCAAGGAACGCGCGTGGT 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 GCAGAAGTTCTCAGGACACTCAACCGCGCAGCAAGTCAATCAAGATCGTCAACGAGAA 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTGTTTCAGATCTCGGTCGGGAAACCCCGCGACTGTCTACTGGCCAAACCCACCGACC 303
QY 103 ValValMetMetValClyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTCATCATGCTGGCAGGTCTCCAGGGTGCAGGTAAAGACCCTCGCAGGTAACTGTCC 363
QY 123 LeuLeuMetArgLysTyAsnLysLysProMetLeuValAlaAlaAspIleTyArg 142
Db 364 AAGCACTCGTCAAGCAG---GGTCACACTCTCTTCTGTTGCTGTGACCTTCACGCT 420
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTy--- 161
Db 421 CCAGGCGCAGTTTCAGCAGCTGCAAAATTTGGGTGAACGCGCAGGCGTACCACTTCGCA 480
QY 162 -----SerGluGlyAsp 165
Db 481 CCGGATCCAGGCGCACCATGATCGTCTCGAGCAGAAATGGGCACTCCACGCTGAT 540
```

```
QY 166 GlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeu 185
Db 541 -----CGATCGAGTACGCGCGCAGGTATCGAAGACCGCAGCGCACCCAGCAC 591

QY 186 AspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlu 205
Db 592 GACATCGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGG 651

QY 206 LeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMet 225
Db 652 GCACGCAACATCCGCGAAGCCATCAACACCTGATGAGTGTCTTTCTCATTCATCCATG 711

QY 226 ThrGlyGlnAspAlaValAlaGluSerPheAspAspGlnLeuAspValThrGly 245
Db 712 ATGTGTCAAGCGCTAGACACCGCGAGAGATTCGCGACCGCGTGCACCTCACCGGT 771

QY 246 ValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSerIleArgSer 265
Db 772 GTTGTCTTGACCAAGCTTGATGCGACGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 831

QY 266 ValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeu 285
Db 832 GTCACCGCGAAGCCATCAAGTTTGCCTCCACTGCTGAAACACTCCAGCACTTCGACGTC 891

QY 286 PheHisProGlnArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIle 305
Db 892 TTCCACCCAGAGCGCATGCGCGAGATTCCTGGCGATGGTGACGTGTCATCATCATC 951

QY 306 GluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuLysLysMetArg 325
Db 952 GAGCAGCGCGAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011

QY 326 GluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGlnValLysAsnLeuGly 345
Db 1012 TCCGCGAGCTCACCTCGGAAGACTTCTTGCACCAAAATGCTGATGATGATGATGATG 1071

QY 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
Db 1072 CCAATCGCAACATCCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1131

QY 366 LeuAsnMetSerGluLysGlnIleLeuAspHisIleLysAlaIleLeuSerMetThrPro 385
Db 1132 ATG---GTTGATGAGAGCAACTCGACGCGATCCAGCGCATTCGCGGCTATGACCCG 1188

QY 386 AlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArgIleAlaLysGly 405
Db 1189 GCGGAGCGGATATCCAAAGATCTTCACGCTTCCAGGCGCAGCGCATCCGCCAACGGT 1248

QY 406 SerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
Db 1249 TCCGCTGTGACCGTGTCCGAAGTAAACAAACTGTGTGAAGCTTCTTCGAGGCTCGCA 1308

QY 426 MetMetLysGlnPheThrGly----- 432
Db 1309 ATGATGGGTCAAAATGGCTGGCATGGTCTCGATCCCGCAGCTGCACACCAAG 1368

QY 433 ---GlyGlyLysGlyLysLysLysArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1369 AGCAGGCGGAGGCGCGGAGGATAGAACGCGCAGCGTAAACACGCGAAGAGGCGCCA 1428

QY 452 AsnLeuPro 454
Db 1429 ACCAGCCA 1437

RESULT 12
US-09-815-242-6166
; Sequence 6166, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6166
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; LOCATION: (1)...(1362)
US-09-815-242-6166

Alignment Scores:
Pred. No.: 1,11e-121 Length: 1362
Score: 1164.50 Matches: 231
Percent Similarity: 71.74% Conservative: 94
Best Local Similarity: 50.99% Mismatches: 121
Query Match: 51.21% Indels: 7
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-6166 (1-1362)

QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGATAATTAAACGATCGTTTTCGCGCAGCTGCGCAATATCATCGCTGAGGAGGAGG 63

QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTCACCTGAAGACACGTAAGAATACGCTGCGGAGTGCATGCGCTGCTGGAGGCG 123

QY 43 AspValAsnPheLysValValLysGluPheLysThrValSerGluArgAlaLeuGly 62
Db 124 GACGTAGCTCTCGCGTAGTGTGCTGATTTATCAATCGCTGAGGAGGAGGAGGAGGAG 183

QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 CATGAAGTTAATAGAGCCCTGACGCGGCGGAGGAGTTCGTCATAATAGTCCGTACCGAA 243

QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTGGTTGCGCGATGGCGGAGAGACACGACCTGACCTGACCTGCGCAACCGCTCGG 303

QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTCGTACTGATGCGGCGCTGCAAGGTGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 363

QY 123 LeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIleThrArg 142
Db 364 AAGTTCTTCCGCGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
```

```
QY 143 ProLaalaleasnGlnLeuGlnThrValGlyLysGlnIleaspIleProValTyrSer 162
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 424 CCGCGCAATCAACAGCTTCAGACGCTGCAGACGAGTGGCGGTTCATTCTCCCT 483

QY 163 GluGlyaspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 484 TCTGATGTTGTCAGAAGCGGTAGATATCGTTTAACGCGCGTGAAGAAGCCAACTG 543

QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspLysAlaLeu 202
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 544 AAAATCAGACGCTGCTGCTGTGTGTGATACCCCTGCTGCTGACGCTGACAGAGCGATG 603

QY 203 MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValVal 222
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 604 ATGGACGAGATCAACAAGTCCATGCTGCTGATTAACCGGTTCAACCCCTGTTGTGGTT 663

QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGlnIleSerPheAspGlnLeuAsp 242
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 664 GACGCCATGACCGGTGAGTACGCGCAATACGCCAAAAGCATTCATGAAGCGGTACCG 723

QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSer 262
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 724 CTTACCGGTGATGTTGACCAAGTACGCGCGATGCCGCGCGGTGCGCGCTCTCT 783

QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 784 ATTGTCACATCACTGGCAACCGATCAAGTTCCTCGGTGTTGGCAGAGAACACTGAGCG 843

QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 844 CTGGAGCGGTTCATCCGACCGCATCGCGTCGCGTATTCCTCGGCATGGCGAGCTACTG 903

QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuLys 322
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 904 TCGCTGATCAAGATATCAAGCAAGATTGACCGCGCGAGCAGAGAAATATGACCGAGC 963

QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluGluLeuAspGlnVal 341
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 964 AAGCTGAAAAAGGTGACGGCTTCGATCTCAACGACTTCTTGAGCAGCTGCGCCAGATG 1023

QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1024 AAAAATATGGCGGCATGCTAGTCTGATGGCAAGCTGCGCGCATGGCGCAGATCCCG 1083

QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIleGln 381
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1084 GATACTCAAGTCACAGATGACGATAAAGTGTGTCGCTGCGATGGAAGCCATCATCAAC 1143

QY 382 SerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArg 401
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1144 TCGATGACGATGAAGAGCGCCTAAGCCAGAAATCATCAAGGTTCCGCTAAGCCGCT 1203

QY 402 IleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPheAsn 421
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1204 ATTGCTCCGGTTGCGGTATGCAAGTGCAGGTCAGGACGCTTAACCGCTCTCTGAAACAGTTCGAC 1263

QY 422 AspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysArg 441
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1264 GACATGACGCGCATGATGAAG-----AAATGAAGAGGCGGGAATG 1305

QY 442 AsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
...  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1306 CGGAGATGATGAGAGCATGAGGGTATGATGCCCCCA 1344
```

RESULT 13

```
US-09-815-242-1949/c
; Sequence 1949, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1949
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-815-242-1949
```

Alignment Scores:

Pred. No.:	2,4e-121	Length:	721
Score:	1158.00	Matches:	234
Percent Similarity:	99.58%	Conservative:	2
Best Local Similarity:	98.73%	Mismatches:	1
Query Match:	50.92%	Indels:	0
DB:	10	Gaps:	0

US-09-943-108a-2 (1-455) x US-09-815-242-1949 (1-721)

```
QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 713 ATGGCATTTGAAGGTTATCAGACGCTTGCACGCGATGCAAAAAATCGTGGTAAG 654

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 653 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAGATTAGCGTTACTT 594

QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 593 GAGGCTGACGTAACTTTAAGTGTAAAGAATTTATTAAACAGTATCAGACGCGCA 534

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 533 TTAGTTCCGATGTAATGCAATCATTAACACAGCGCAACAAGTTATTAATAAGTTCAA 474

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 473 GATGAATTAACGCGAGTTGATGGTGGAGAAAATACGTCGATTAATATGTCAAATAACCA 414

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 413 CCTACTGTTGTTATGATGGTTGGTTTACAGGTGCTGTAAACACACACACTGCAGGTAAA 354

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 TTAGCATTATGATGCGTAAAAAATACACAAAAAACCTATGTTAGTTCGACAGATATT 294

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 293 TATCGTCACGACGCGATAAATCAATTACAAACAGTAGGGAAACAAATGATTATTCCTGTA 234
```

```
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||
Db 233 TACAGTGAAGGAGATCAAGTAAAGCCACAAATTTGAATTAATGATTAAACATGCT 174

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||
Db 173 AANGAAGAACATTGACTTTGTAATCATGTATACAGAGTTCGATTACACATCCGATGAA 114

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||
Db 113 GCATTGATGAAGATTAAGAGAGTAAGAGACATGCTAAACCAACAGTAATATGTA 54

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPhe 237
|||||
Db 53 GTTGTGATTCATTAATGACGGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTT 3
```

RESULT 14

```
US-09-815-242-7850
; Sequence 7850, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7850
; LENGTH: 1374
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1374)
US-09-815-242-7850
```

Alignment Scores:

```
Pred. No.: 4,79e-120 Length: 1374
Score: 1150.00 Matches: 231
Percent Similarity: 70.90% Conservative: 93
Best Local Similarity: 50.55% Mismatches: 121
Query Match: 50.57% Indels: 12
DB: 10 Gaps: 3
```

US-09-943-108a-2 (1-455) x US-09-815-242-7850 (1-1374)

```
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
|||||
Db 4 TTGCAAAACCTTACAGACCGGCTCTCGCAGACGCTTCGCGCATGTACCGGCAAGGCCAAG 63
```

```
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
|||||
Db 64 CTACCGAGGACACATCAAGGACACTTCGCGGAAGTGGCATGGCCCTGCTCGAGGCC 123

QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
|||||
Db 124 GAGCTGGCCCTCGCGGTGGTCAAGGACTTCGTCAACAAGGTCAAGAGCGCGCGTGGT 183

QY 63 SerAspValMetLysSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
|||||
Db 184 ACCGAGGTCTCGAGAGCGTGACCCCGGACAGCGCTTCGTGAAGATCGTCGCCGCCGAG 243

QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProPzofthr 102
|||||
Db 244 CTGAGGAGCTGATGGGGCGGCCAACAGAGGACCTGGCGCTGAGCGTCGCGCGCGGGG 303

QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
|||||
Db 304 GTGATCCTGATGGCGGCTGCAAGGGCGGGCAAGACACCCCGGGCAAGCTGGGCG 363

QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
|||||
Db 364 CGCTTCCTTAAGGAACCCAGAGAAGTCGGTGATGGTGGTTCGCCGACGCTACCGA 423

QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
|||||
Db 424 CCGCGCGCTATCAAGCAGCTGGAACCCCTGGCGGGCGAGGTGCGCGTAACCTTCTCCT 483

QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
|||||
Db 484 TCGAGGTCTAGCAGCAAGCGGTCGCCATCGCGGAGCGCGGATCCCGGAGCGGAGCTG 543

QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
|||||
Db 544 AATTCATCGAGTGGTGTGATCGTCGATACCGCAGCGCGCTGCACATCGATGCCGACATG 603

QY 203 MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValVal 222
|||||
Db 604 ATGGAGAGATCAAGCAGGTGCACGGCGGCGATCAAGCGCGGGAACCCCTGTTCTGCTC 663

QY 223 AspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPheAspAspGlnLeuasp 242
|||||
Db 664 GATGCCATGACCGCGGATGGATGGCCCAACCGCCAGGCTTCATATGACGCTTGGCGG 723

QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
|||||
Db 724 CTGACCGGGGTGCTCTCACCAAGGTGCGCGGACGCGCGTGGTGGTGGCGGCTCTCG 783

QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
|||||
Db 784 GTCGGGCGATCACCGCAAGCGGATCAAGTTCTCTGGCATGGCGGCAAGAGCGGAAGCG 843

QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
|||||
Db 844 CTGATCCGTTCCATCCCGACCGCGTGGCTCGCATCCCTCGGCATGGCGGACGCTGCTC 903

QY 303 SerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeuGluLys 322
|||||
Db 904 AGCTTGATCGAAGCGCGAGCAGACCTCGACCGCGCAAGAGCGGACGAGCTGGCGAAG 963

QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnVal 341
|||||
Db 964 AAGATCAGAGAGGCAAGGGCTTCGACCTCGAAGAGCTTCGCGGACCATGTTGCAACAGATG 1023

QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIlePro-----GlyMetAsn 358
|||||
Db 1024 AAGACATGGCGGCGCTCGCGGGGCTCATGGACAAAGCTCGCGGATGCTCGCGGGGTCAAC 1083

QY 359 LysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378
|||||
Db 1084 CTGGCCGATGGGCAATGCCCCAGGTGCTCGGAGAAACAGTTCGAAGCAGATGGAGGCG 1143

QY 379 IleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArg 398
```

```

Db 1144 ATCATCAACTCGATAGCGCGCGGACGCGGATCCGGAATGATCAGCGGCTCGCG 1203
QY 399 LysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLys 418
Db 1204 AAGCGCGCATCGCTTGGCTCGGTACCCAGGTGCAGGAGCGTCGCGCGCTGATCAAG 1263
QY 419 GlnPheAsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysLysLys 438
Db 1264 CAGCACAGCAGATGCAGAGATGATGAGAGAGTCAACGCCAAGGCG----- 1311
QY 439 GlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1312 -----GGCATGCCCAAGATGATGCGAGCATGGCGATGTTTC 1350

```

RESULT 15

```

US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamanoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868

```

```

Alignment Scores:
Pred. No.: 1,09e-118 Length: 1389
Score: 1138.00 Matches: 225
Percent Similarity: 70.67% Conservative: 93
Best Local Similarity: 50.00% Mismatches: 122
Query Match: 50.04% Indels: 10
DB: 10 Gaps: 2

```

```
US-09-943-108a-2 (1-455) x US-09-815-242-6868 (1-1389)
```

```

QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTTGAGATTATTCGGATCGCTTCCAAAACCTAGCTAATATACAGGAAAGGCGGT 63
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42

```

```

Db 64 TTAACGGAAGATAATTAAGAAACCTTACGCGAAGTCGCTATGCATTAATTGAGGC 123
QY 43 AspValAsnPhelysValValysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GATGTTCCCTTGCCTGTGTCGCTGAATTTATGCGAAAAAGTAAAAAGAGCGCTGGG 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 GAAGAGTCAATAAAGATTTACGCCAGGCGAAGATTTCTTAAATGCTTACGCTGAG 243
QY 83 LeuThrLysLeuMetGlyGlyGlnAsnThrSerIleAsnMetSerAsnLysProProThr 102
Db 244 CTTGAAAAACCATGGCGAAGCATGAGAGTTTAAATCTCGCAACCCACCACGACGA 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTTATCTTAATGGCGGTTTACAAGGGGCGGTAAACCCACGAGTGGGTAAATTGGCA 363
QY 123 LeuLeuMetArgLysLysTyraLysProMetLeuValAlaAlaAspIleTyraG 142
Db 364 AAATCTTGGGTGAGCGCATTAAGAAAGAGTGTAGTGTGTCGTGACGATATCGC 423
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySer 162
Db 424 CCTGCTCGGATTAAGCAACTTGAACCTTGGCTCAATCCGTTGGCGTGGATTTTTC 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
Db 484 TCGGATGTTAAACAAACCCCGTTGATATTGCTAAATCGCGCTTGCCTGCTCAAAA 543
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
Db 544 AAATCTACGATGTTGATGTGGATACGCGAGTCCGCTACACGTTGATACAGAATG 603
QY 203 MetAsnGluLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetLeuVal 222
Db 604 ATGGACGAAATCAAGCAAGTCATTCGATTAATCAATCAATCAAACTCTTTTCACTG 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGluLeuAsp 242
Db 664 GATCGATGATCGTCAAGATCGCGCAAAATCAAGCAAAACGCTTTAATGAGCATTC 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTTACAGGGGTTATTTGACGAAAGTGCAGCGTGATCGCGCGGTGGTGGCGCTTATCG 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 784 ATTGCTCAAAATCACAGGTAAACCAATCAAAATCTTGGGTGTGGCGCGAAGACAGAG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTTGAGCCATTCCTCTGATCGTGCTCTCCGCTATTTTGGGATGGGCGATGTCCT 903
QY 303 SerLeuIleGluLysAlaGlnAspValAspGlnLysAlaLysAspLeuLys 322
Db 904 TCCCTTATCGAAGATCTTGACGTTCTGTTGATCGTGAAGAGGAAAAAATGGCGCAG 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGluLeuAspGlnVal 341
Db 964 AAATTCAGAAAGGCGGATGATTTTACTTATGATGATTTCCGCGAAGACGCTGATCGA 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
Db 1024 AAAAAATGGCGCGCATGATGCTATGCTTCAAGAAATTTACAGGTCGCAAAAATTTGCT 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleGln 381
Db 1084 GAGCACCTTAAATCAAGTAGATGACAAATGTTTGTCAAAATGGAAGCGATCAATTAC 1143
QY 382 SerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLysArg 401

```

```

Db 1144 TCCATGACCCCTAAAGAACGTGCCAACCCAGATATTATCAAGGATCTGCGCTCGTGT 1203
QY 402 IleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsn 421
Db 1204 ATTCATTAGGCTCTGGCACTCAAGTGCAGATGTCATTAATTAATTAATTAATTAATTAAT 1263
QY 422 AspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLysArg 441
Db 1264 GAAATGCAACGTATGATGATGAAGAAATGCGTAAGCGGCGC----- 1302
QY 442 AsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1303 -----ATGGCTAAATATGATCGTGGATG 1326

```

Search completed: February 25, 2003, 04:03:39
Job time : 113 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 01:10:37 ; Search time 1563 Seconds
(without alignments)
4714.617 Million cell updates/sec

Title: US-09-943-108a-2
Perfect score: 2274
Sequence: 1 MAFEGLSERLQATMKMRGK.....GKKGRNQNMKLKGMNLPF 455

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=framet_p2n.model -DEV=slp
-Q/cgcn2_i/USPTO_spool/US09943108/runat_21022003_151631_1824/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943108 -CGCN_1_1_2874 @runat_21022003_151631_1824 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	1074	47.2	2159	17	BH770912 LIMGtag64
c 2	757.5	33.3	875	17	AF075878 AF075878
c 3	715.5	31.5	773	13	B1934118 EST554007
c 4	631	27.7	656	14	B0805434 WHE3566.H
c 5	630	27.7	711	12	BF940797 7099A06.X
c 6	616.5	27.1	2387	11	AK011928 Mus muscu
c 7	594.5	26.1	688	17	BH394904 AG-ND-138
c 8	562.5	24.7	605	10	BE432304 EST398833
c 9	558	24.5	593	10	AW648731 EST327101
c 10	555.5	24.4	581	9	AI977909 496035E06
c 11	534	23.5	677	13	BM413418 EST587745
c 12	520.5	22.9	1200	11	AY109321 Zea mays
c 13	513.5	22.6	533	10	AW930742 EST356585
c 14	506.5	22.3	677	14	BQ996389 QGG12L07
c 15	485	21.3	514	13	BM323434 PIC1_19.C
c 16	479	21.1	1020	17	CNS07B0K T3 end of
c 17	473	20.8	791	17	AQ327477 nbxb0041K
c 18	465.5	20.5	1032	11	AY106667 Zea mays
c 19	463.5	20.4	483	10	BE596266 P11_51.F0
c 20	451	19.8	777	17	BH402176 AG-ND-122
c 21	448.5	19.7	490	13	BM325805 PIC1_53.A
c 22	438.5	19.3	554	17	BH369841 AG-ND-126
c 23	437.5	19.2	481	17	BH376496 AG-ND-178
c 24	433.5	19.1	880	17	CNS077VW T7 end of
c 25	420	18.5	465	9	AI087737 SWACAC23
c 26	419	18.4	476	10	AW033339 EST276910
c 27	419	18.4	899	14	BQ736950 AGENCOURT
c 28	414.5	18.2	547	10	BE461473 EST412892
c 29	412.5	18.1	688	13	BI933601 EST553490
c 30	412.5	18.1	1108	9	AU091185 AU091185
c 31	408	17.9	1059	14	BM806783 AGENCOURT
c 32	405	17.8	1012	17	AF094895 AF094895
c 33	400	17.6	949	17	CNS076TA T7 end of
c 34	398.5	17.5	814	17	AZ529841 ENTPC084TR
c 35	392.5	17.3	889	17	AZ682189 ENTPC084TR
c 36	389.5	17.1	773	13	BI454259 603170687
c 37	388	17.1	761	9	AL655153 AL655153
c 38	385.5	17.0	656	12	BF641333 NF064FI2I
c 39	380.5	16.7	694	13	BJ473333 BJ473333
c 40	378.5	16.6	874	12	BG679799 602626480
c 41	377	16.6	707	17	AQ654845 Sheared D
c 42	375.5	16.5	753	13	BG921695 602825514
c 43	375	16.5	382	10	AM649149 EST327603
c 44	374.5	16.5	975	13	BG919869 602834428
c 45	369.5	16.2	573	12	BG317280 547027E11

ALIGNMENTS

RESULT 1
BH770912/c
LOCUS BH770912 2169 bp DNA linear GSS 01-MAY-2002
DEFINITION LIMGtag644 MG1363 Random Sequence Tag Library Lactococcus lactis
subsp. cremoris genomic, DNA sequence.
ACCESSION BH770912 GI:20373869
VERSION BH770912.1
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris
ORGANISM Lactococcus lactis subsp. cremoris
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 2169)
AUTHORS Bolotin,A., Ehrlich,S.D. and Sorokin,A.
TITLE Studies of genomes of dairy bacteria Lactococcus lactis

JOURNAL Sci. Aliments, (2002) In press
 COMMENT Contact: Sorokin A
 Genetique Microbienne
 INRA
 CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
 Tel: 33 1 34 65 25 16
 Fax: 33 1 34 65 25 21
 Email: sorokine@jouy.inra.fr
 best homologue in strain IL1403 is fth (91%)
 Class: shotgun
 High quality sequence start: 30
 High quality sequence stop: 2141.
 High quality sequence stop: 2141.
 FEATURES
 source
 1..2169
 /organism="Lactococcus lactis subsp. cremoris"
 /strain="MG1363"
 /db_xref="taxon:1359"
 /clone_lib="MG1363 Random Sequence Tag Library"
 /note="Vector: pSGM2; Site_1: SmaI; Library of
 chromosomal fragments of L.lactis strain MG1363 was
 prepared by partial AluI digestion or by sonication."
 BASE COUNT 642 a 429 c 380 g 718 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.02e-108 Length: 2169
 Score: 1074.00 Matches: 218
 Percent Similarity: 73.86% Conservative: 73
 Best Local Similarity: 55.33% Mismatches: 82
 Query Match: 47.23% Indels: 21
 DB: 17 Gaps: 2
 US-09-943-108A-2 (1-455) x BH770912 (1-2169)
 QY 76 IleIysIleValIClAspGluLeuThrLysLeuMetGlyGluAsnThrSerIleasn 95
 Db 2168 ATCAAAATTTGGTATGAGAAATTAACAACATCTTATAGCGGTGGTGAAGCCGAAATTAATCT 2109
 QY 96 MetSerAsnIysProProThrValValMetMet-ValGlyLeuGlnGlyValAlaGlyLysTh 115
 Db 2108 AAATCTCCTTAAATCCACACATTAATCATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2049
 QY 115 rThrThrAlaGlyLysLeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLe 135
 Db 2048 AACCTTTGGTGTAAACTGCAAAACCTTAAAGAAACAAATGCTGCCACATGAT 1989
 QY 135 uVal-AlaAlaaspIleYrArgProAlaAlaIle-AsnGlnLeuGlnThrValGlyLys 154
 Db 1988 GATTGGCGCGGCGGTTATGCTCAGCAGCGATGGACCAATTAACAACTTGGGAGAA 1929
 QY 155 GlnIleAspIleProValItyr-SerGluGlyAspGlnValIysProGlnGlnIleValThr 174
 Db 1928 CAATTCGACATTCCTGTTATGATGAGGAAACGCTGAAACACCCGTTAATATTGTCGT 1869
 QY 175 AsnAlaLeuLysHisAlaLysGluCluHisLeuAspPheValIleAspThrAlaGly 194
 Db 1868 AATGGCTCTTTAAAGACTCAAGAAGACGTAAAGATATGCTGATTGATACGCGAGGT 1809
 QY 195 ArgLeuHisIleAspGluAlaLeuMetAsnCluLeuLysGluValIysGluIleAlaLys 214
 Db 1808 CGTCTTGAAATGATGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1749
 QY 215 ProAsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValala 234
 Db 1748 CCACAGAAATTTTACTCGTCTGATGCGATGCTGCTCAAGTGGCCGACAGATGGCT 1689
 QY 235 GluSerPheAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAsp 254
 Db 1688 AAAACCTTTGATGAAATTAATGATATACCGGTGTTATCATCAACCAAACTCGATGGGAC 1629
 QY 255 ThrArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleIysPheVal 274
 Db 1628 ACACGTGGTGGCGCTTTATCAATTCGTGAATTAACCGGAAACCTTATTAATTAATTAAC 1569

QY 275 GlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArg 294
 Db 1568 GGGACTGGTGAATAATTAACGACCTTGAATTTTACCCCTGATCGTATGAGCTCAAGA 1509
 QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnAspValAspGln 314
 Db 1508 ATTCTTGGTATGGGTATGATGCTCACTCTGATTGAGAGGCGCAGCTAATATGATGAG 1449
 QY 315 GluLysAlaLysAspLeuGluLysLysMetArgLysSerPheThrLeuAspAspPhe 334
 Db 1448 GACAATCTGCAGAACTCGTGAATAAATGGCTGAACCCGTTTGTATTGATGAGGATTC 1389
 QY 335 LeuGluLeuAspGlnValIysAsnLeuGlyProLeuAspIleMetLysMetIle 354
 Db 1388 GTTGACCACTAGATCAAGTTACAATAATGGGACCAATGGAGATATATGAAAAATGAT 1329
 QY 355 ProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAsp 374
 Db 1328 CCAGGAATGTCAGAAATGCTGGCTTGGATAAAAGTTAAAGTTGACCCCTAAAGATGTTGCT 1269
 QY 375 HisIleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeu 394
 Db 1268 AGAAACGAGCAATGATGTCAATGACCCAGCTGAACGTCATTAGAACGACAGACTA 1209
 QY 395 AsnValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsn 414
 Db 1208 TCGCAGCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1149
 QY 415 ArgLeuMetLysGlnPheAsn-----Asp 422
 Db 1148 AAATTCATTAAAGCAATTTAATCAATCCAAAGAAATGATCAAGGAATCATGAATGGTGAT 1089
 QY 423 MetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLysArgAsn 442
 Db 1088 ATGAATCGGATGATGCAAAATGATGGTGGAGCTGGTGGG----- 1047
 QY 443 GlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
 Db 1046 CAATGCCAATATGCCACCGGTCCTGGAATGCCA 1011
 RESULT 2
 AF075878/c 875 bp DNA linear GSS 29-AUG-2000
 LOCUS AF075878 Salmonella typhimurium LT2, Lambda DASH II Salmonella
 DEFINITION typhimurium genomic clone 245-T3, DNA sequence.
 ACCESSION AF075878
 VERSION AF075878.1 GI:3320748
 KEYWORDS GSS.
 SOURCE Salmonella typhimurium.
 ORGANISM Salmonella typhimurium
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Salmonella.
 REFERENCE 1 (bases 1 to 875)
 AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
 TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
 comparison to the Escherichia coli K12 genome
 JOURNAL FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
 MEDLINE 95243757
 COMMENT Contact: McClelland M
 Molecular Biology
 Sidney Kimmel Cancer Center
 3099 Science Park Road, San Diego, CA 92121, USA
 Email: mclelland@lifesci.sdsu.edu
 Class: shotgun.
 Location/Qualifiers
 source
 1..875
 /organism="Salmonella typhimurium"
 /strain="LT2"
 /db_xref="taxon:602"
 /clone="245-T3"
 /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
 /note="Vector: Lambda DASH II; sequenced using Li-Cor

```

BASE COUNT      178 a   266 c   217 g   214 t
ORIGIN
sequencer"
Alignment Scores:
Pred. No.:      1,51e-73      Length:      875
Score:          757.50      Matches:      152
Percent Similarity: 73.08%      Conservative: 57
Best Local Similarity: 53.15%      Mismatches:  76
Query Match:    33.31%      Indels:      2
DB:             17      Gaps:      1
US-09-943-108A-2 (1-455) x AF075878 (1-875)
QY 87 MetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThrValValMetMet 106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 867 ATGGCGAAGACCAAGAGCGTGAATTTTAC-TGCGACGCGCCAGCGTAGTATGATG 809
QY 107 ValGlyLeuGlnGlyValGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArg 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 808 GCGGCTGCGAGGGGGGGTAAACACACAGCGTCGGTAGTGGGTAAATCTCTGGC 749
QY 127 LysLysTyAsnLysLysProMetLeuValAlaAlaAspIleTyArgProAlaAlaLe 146
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 748 GAGAAGCACAAGAGAGAGTCTGCTGCTGCGGACGCTCTATCGCGGGGGCGATC 689
QY 147 AsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySerGlnGlyaspGln 166
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 AAACAGCTCGAAGCGTGGCTGACGAGGTGGCGGTATTTCTCCCGTCTGATGTCGCG 629
QY 167 ValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuasp 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 CAGAAACCGGTGATGCTGCTCAGCGCGCGCTGAAGAGAGCGAAGCTCAATCTACGAC 569
QY 187 PheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeu 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 GTGCTGCTGGTGATACGCGCGGTCTGCTGATGTTGACGAGCGATGAGGACGAATC 509
QY 207 LysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValaspSerMetThr 226
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 AAACAGGTCCAGCTTCTATCAGCCAGTAGAAGCGTGTGTGTCGTCGATGCGATGACC 449
QY 227 GlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyVal 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 GGTGCGATGCGCGGAATACCGCAAAAGCCTTTACGAGAGCGCTGCCGTTAACCGCGTG 389
QY 247 ThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAlaLeuSerIleArgSerVal 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 GTGCTGACCAAGTTGATGGTGATGCGCGGTGGCGGTGGCGCTCTCTATTCTCATATC 329
QY 267 ThrGlnLysProIleIlePheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPhe 286
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 ACCGGCAAGCCGATTAAATTCCTCGGTGTCGGCGAGAAACCGACGCGCTGGAGCCATTC 269
QY 287 HisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGlu 306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 CATCCGATCGATCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
QY 307 LysAlaGlnGlnAspValaspGlnGlyAlaLysAspLeuLysLysMetArgGlu 326
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 GATATCAAGCAAGTGTGCGCGCACAGGCTGAGAGCTGGCGACCACTGAAGAAA 149
QY 327 ---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 GCGGACGTTTCGACCTGAACGACTCTCTGGAACAGCTCAACACAGATGAATAACATGGG 89
QY 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GGTATGCCACGCTGATGGGCAAAATTACCGGGCATGGGCGCATGGGCGATATCCGGCAACAGTTAAA 29
QY 366 LeuAsnMetSerGlnLys 371
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 28 TCGCAGATGGATGACAAAG 11

```

RESULT 3

BI934118

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..773

/organism="Lycopersicon esculentum"

/cultivar="RA496"

/db_xref="taxon:4081"

/clone="CF0D18J16"

/clone_lib="tomato flower, anthesis"

/tissue_type="flower"

/dev_stage="anthesis"

/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Cornell University; sequencing: The

Institute for Genomic Research; Flower buds and flowers

were taken from greenhouse plants (4-8 wks old, RA496).

They were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 227 a 129 c 210 g 207 t

ORIGIN

Alignment scores:

Pred. No.: 5,83e-69 Length: 773

Score: 715.50 Matches: 134

Percent Similarity: 75.19% Conservative: 60

Best Local Similarity: 51.94% Mismatches: 63

Query Match: 31.46% Indels: 1

DB: 13 Gaps: 1

US-09-943-108A-2 (1-455) x BI934118 (1-773)

QY 39 LeuPheGluAlaAspValAsnPheIleLysGluPheIleLysThrValSerGlu 58

Db 1 CTTTATGAAGCTGATGTAGTCTCCCAAGGTTTGTTCAGTCTGTAGTGA 60

QY 59 ArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIle 78

Db 61 GAAGCCGTTGGCGAGTGGTGTGATTCGAGGTGTAACACAGATCAGCACTAGTAACT 120

QY 79 ValGlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsn 98

Db 121 GTAGTGCACGACCTGTGAACCTGATGGGTGGAGAGGTTTCTGAACCTGCTTTGCTAAA 180

QY 99 LysProProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAla 118

Db 181 TCTAAGCCACCCGTAATACTATTGGCGGCTCTACAAGGTGTTGGAGACAACTGTAGT 240

```

QY 119 GlyLysLeuAlaLeuMetArgLysLysTyrAspLysProMetLeuValAlaAla 138
Db 241 GCRAAGTACCTTTATCTA---AAGACAGAGGTAAAGTGCATGCTGCTGGGA 297
QY 139 AspIleYrArgProAlaAlaIleAsnGlnLeuThrValcLysGlnIleAspIle 158
Db 298 GACGTGTACAGACCTGCTATTGACCACTTGTTATTTGGTAAACAGGTTGATGA 357
QY 159 ProValTySereGluGlyAspGlnValLysProGlnGluIleValThrAsnAlaLeuLys 178
Db 358 CCTTTTATGACAGACAGACAGATGTAACACCTGCAGAAATAGCCCGAAGGATTACAA 417
QY 179 HisAlaLysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIle 198
Db 418 GAGCCAAAAGAGAGATCTAGTCTATATGATGATACAGCTGGAGACTTCAGATA 477
QY 199 AspGluAlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGluIle 218
Db 478 GATAAAACTATGATGATGAATTAAGAGAGCTGAACGGGTACTGAACCCACAGAGTT 537
QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 538 TTACTTTTGTGATGCAATGACTGCGCAGAGAGCTGCACCTTTGTCACACATTCAT 597
QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 598 CTCGAATTTGGGATTACTGTCCTGCTGACGAGCTAGATGAGATCTAGGGGTGGA 657
QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlu 278
Db 658 GCACCTTTTGTCTCAGAGAGGTATCAGGAAGCCAAATCAAGTCTGTAGGAAGGGTGA 717
QY 279 LysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeu 296
Db 718 CGTATGAGGACCTTGAACCTTTCTAUCCTGACCGCATGCTGGACGTATTTTA 771

RESULT 4
BQ0805434
LOCUS
DEFINITION WHE3566_H12_P24ZS Wheat developing grains cdna library Triticum
aestivum cDNA clone WHE3566_H12_P24, mRNA sequence.
ACCESSION BQ0805434
VERSION BQ0805434.1 GI:22029643
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 656)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cdna library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
Location/Qualifiers
1..656
/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3566_H12_P24"
/clone_lib="Wheat developing grains cdna library"

```

```

/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/Note="vector: Lambda ZAP II, excised phagemid; Site_1:
EcoRI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24
, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16
, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16
, 20, 24, 28, 30 DPA and total RNA was prepared by S.
Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give plasmidscript SK(-) phagemids
in the TJ Close lab (Chin, Close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
BASE COUNT 200 a 115 c 179 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 11e-59 Length: 656
Score: 631.00 Matches: 116
Percent Similarity: 75.23% Conservative: 48
Best Local Similarity: 53.21% Mismatches: 54
Query Match: 27.75% Indels: 0
DB: 14 Gaps: 0
US-09-943-108A-2 (1-455) x BQ0805434 (1-656)
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 3 TACAGGCTCTGCCATTCATCTACTCTGCTGGTGACAGGTGGGTGCCAGTT 62
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 63 TACTCAGAGGAACCGCGCCCAACCTGCAGAAATACCAAGAATCCCGTGAAGAGCG 122
QY 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 123 AAAAGAAATATATATGATCGCATGCTGTGATGATCTGCTGCAGATCGCATGATAAA 182
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetIleu 220
Db 183 ACAATGATGTTGAATGTAAGAAAGATGAAGAGCCAGTTCATCTACAGAGATTTCGTT 242
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 243 GTTGTGATGCTGACTGCTGCCAGGAAGCTGCAGCATTTGTCACAACTTCAATATTGAA 302
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 303 ATTGTATATCCGGTCTATATTGATTAATGATGATTCAGGGGTGAGACGCA 362
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysIleu 280
Db 363 CTTAGTGTGAAGAGGTGTCTGAAGAGCCCATCAAGTTTGTGGCGGAGAGAGCAATG 422
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300

```

```

Db 423 GAGGACCTTGAGCTTTCATCCGACCGCATGCGCACACAGTGTGTTGGATGGGAGAT 482
    ::: |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysLeu 320
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 483 GTCCCTTATGTTGTAAGAAACACAAAGATGCTGCGCAAGAGATACCATGGAACTG 542
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
    ::: |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 543 CAGAAGAAGATCAGAGTGGGAAATTTGACTTCATGACATTTTAAAGCACACAAAT 602
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 341 VallysAsnLeuGlyProLeuAspPheLeuMetLysMetLysMetLysMetLys 358
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 603 GTTGGGAAAGGGGATCCATGAGCCGTGTAATCGGATGATCCAGGATGAAC 656
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 5
BF940797/c
LOCUS
DEFINITION
IMAGE:3280811_3' similar to SW:SR54_BACSU_P37105 SIGNAL RECOGNITION
PARTICLE PROTEIN ?; mRNA sequence.
ACCESSION
BF940797
VERSION
BF940797.1 GI:12358117
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 711)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. James Lupski
cDNA Library Preparation: Lupski Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LILNL at:
info@image.lilnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 433.
FEATURES
    Location/Qualifiers
        1..711
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3280811"
            /clone_lib="Lupski_dorsal_root_ganglion"
            /sex="male"
            /tissue_type="dorsal root ganglia"
            /dev_stage="adult, 36 yr"
            /lab_host="DE10B"
        /notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
        NotI; Site_2: SalI; cDNA made by oligo-dT priming.
        Directionally cloned using the following adaptors:
        5'-TCGACCCAGCGCTGCG-3' and
        5'-GACTAGTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
        1 kb for average insert length 1.7 kb. This is a primary
        library, non-amplified. Library constructed by Life
        Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
        College of Medicine) and is available through Life
        Technologies."
BASE COUNT
193 a 145 c 88 g 283 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1,57e-59 Length: 711
Score: 630.00 Matches: 134
Percent Similarity: 77.45% Conservative: 48
Best Local Similarity: 57.02% Mismatches: 45
Query Match: 27.70% Indels: 8

```

```

Db: 12 Gaps: 3
US-09-943-108A-2 (1-455) x BF940797 (1-711)
Qy 202 LeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeuVal 221
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 708 TTAATGCCAGAGGTAAAGATGTTAAAGAAAGTTTCAAAATCTAGTGAATATTATTAGTA 649
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 222 ValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeu 241
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 648 GTTGATGCTATGATGCGCCACCATGCTCTTAAATGTATGACAGAGACATTAATAATCAATCA 589
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 242 AspValThr-GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaLe 261
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 588 GATCTTTCAGNGAATATATATTACAAAGTTAGATGGTGATACAGAGGTGGTGGTCTCT 529
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 261 uSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerSerGluLysLeuAs 281
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 528 TTCAATAAGACACATTACTGCGAAGCCCAATTAATTTGTTGGTGTGAACAGATGAG 469
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 281 pGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspVa 301
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 468 TCATATTGAATATTCATCCAGATAGATGCTTCAAGAAATATTAGGAATGGGAGATGT 409
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 301 lLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGl 321
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 408 CTTTCTTTTAAATAGAAAAGCTCAGCAAGCTATTGACCAGAGATGACAGCAAGTAAGTAA 349
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 321 uLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGlnVa 341
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 348 TGAATAAATGTTAAATCAAGAATTTAACTTTGATGACTACTTATCAGCTATGATGATCAAT 289
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 341 lLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsn-Lys---M 360
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 288 GAAAAGCTTGGACCTATAATAATGATAGAGATGATCCAGGTGTAAACAAGGAAAC 229
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 360 eLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleI 380
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 228 TTGAGGCAATTGAT---TTTCTCAAGGAGAAAACAAATGGCTACAGTTAAAGCAATCA 172
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 380 leGlnSerMetThrProAlaGluArgAsnProAspThrLeu-----AsnValSerA 398
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 171 TACAATCGATGACAGCTAAAAGAAAGAAAACAACTCTCTTTAGTAATAGGAAATGGTCAA 112
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 398 rGlyLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMet 417
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 111 GAAAGAGAAGATAGCTAAAAGGTTCTGGTACACACAGTACAAGAAAGNTAATAAAGTCTT 52
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 418 LysGln-PheAsnAspMetLysLysMetMetLysGln 429
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 51 AAAAGCTATGAATGATGATAAAAGCAAAATGAAGCAA 15
    |||||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 6
AK011928
LOCUS
DEFINITION
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610209C12:signal recognition particle 54
kDa full insert sequence.
ACCESSION
AK011928
VERSION
AK011928.1 GI:12848353
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2610209C12.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

```

REFERENCE
AUTHORS

2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159

REFERENCE
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,F., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861

REFERENCE
AUTHORS

4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamataka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyokawa,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
PUBMED
11217851

TITLE
JOURNAL
MEDLINE
PUBMED

5 (bases 1 to 2387)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

REFERENCE
AUTHORS

6 (bases 1 to 2387)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

TITLE
JOURNAL

7 (bases 1 to 2387)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT
FEATURES
source

8 (bases 1 to 2387)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

/strain="C57BL/6J"
/db_xref="FANTOM DB:2610209C12"
/db_xref="MGI:1905318"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
1. .2387
/gene="Srp54"
238. .1752
/gene="Srp54"
/notes="data source:MGI, source key:MGI:1346087, evidence:ISS
putative
signal recognition particle 54 kDa"
/codon_start=1
/protein_id="BAB27921.1"
/db_xref="GI:12848354"
/db_xref="MGI:1346087"
/translation="MYLADLRKTKTSALRSLSNATINEEVLNMLKEVCTALLEADY
NIKLVKOLRENVKSAIDLEEMASGLNKRKTIQHAVFKELVLDVPGVKAMPTKRGON
VINEVGLQSGKTTCKSLAYYORQGWKTKLICADTERFAGADOLKONATKARIPY
GSTEYMDPVILASEGVEKEKNEFEIIVDTSGRHOEDSFEEMLOYSNAIOPDNIV
YVNDASIGQACERQAKFKVDVASVIVYKLDGCHAKGCGALSVAIAKSPFIPIGTG
EHIDDFEPTQPFISKLLMGDIEGLIDKNEKLDNEALIEKLEKHGQFTLRDYTE
OFQNIKMGPFQIOLGMITPGFTDFMSKNEQESMARLKKMTIMDSNNDDELSDTG
AKVSPQRIQIQRVARGSGVTRDVBELTQYTKFAQVYKMGGIKGLFGGDMKNV
SOSOMAKLNQAKMMDPRVLHMGMGAGLQSMRMRFQOQAGNMKGMGFNMM"
BASE COUNT 741 a 420 c 560 g 666 t
ORIGIN
Alignment Scores:
Pred. No.: 2,29e-57 Length: 2387
Score: 616.50 Matches: 150
Percent Similarity: 52.12% Conservative: 96
Best Local Similarity: 31.78% Mismatches: 201
Query Match: 27.11% Indels: 25
DB: 11 Gaps: 8
US-09-943-108a-2 (1-455) x AK011928 (1-2387)
Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 238 ATGGTTCTTAGCAGATCTTGGAGAAAAATAACATCAGCATTTACGCTATTGAGCAATGCC 297
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 298 ACCATTATCAATGAAGAGGTATTAATGCTATGCTGAAGAAGTATGTACAGCATTTATG 357
Qy 41 GluAlaAspValAsnPhelLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 358 GAAGCAGATGTTAATATTAAACTAGTAGCACTAGCAAGAAAATGTAAGTCTGCAATT 417
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 418 GATCTTTGAAGAGATGGCATCTGGACTGCAACAAAAGAAAATGATTCAGCATGCTGTTT 477
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 478 AAGAACTTGTAAAGCTTGTAGACCTGGAGTTAAGCGTGAACACCGACTAAGGGAAG 537
Qy 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 538 CAAATGTCATCATCTTTGTTGGATTGCAAGGAGTGGTAAACGACCAACATGTTCAAG 597
Qy 121 LeuAlaLeuMetMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 598 TTACGATATTATTACGAGGAAA--GTTTGAAGACCTGTTTATATGTCGATACA 654
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 655 TTCAGACGAGGAGCGCTTTGACCAGCTAAACAGAAATGCCAACAAAGCAAGATTCGTTTC 714

```

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
||| : : : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 TATGGCAGCTATATCTGAATGGATGCTCTCATCATCTGCTTGAAGAGTGGAGAAATTC 774

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 775 AAAAAAGAAATTTTGAATATATATTTGTTGTATACAACTGGTGTACACAAAGAACG 834

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 835 TCTTTATTTGAAGAAATGCTTCAAGTTTCTAATGCTATACAACTGATACACATTTTAT 894

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 895 GTGATGGATGCATCATCGGACAGCGTTGTGAGGCCGCGGAGAGCGCTTTAAAGACAAA 954

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
: : : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 955 GTAGATGTAGCTTCAGTAAATAGTGACAAAACCTCGACGGTCTATGCGAAAGGAGCGGTGCT 1014

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1015 CTTAGTGCAGTTGCTGCCACAAAAGTCCCAATTTATTTTCATTGTACAGGGGACATATA 1074

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 GATGATTTTGAACCTTTCAAAACACACACCTTCATCAGCAACACTCCTTGAAGTGGGTGT 1134

QY 301 ValLeuSerLeuIleGluLysAlaGlnGln--AspValAspGlnGluLysAlaLysAsp 319
: : : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 ATTGAAGGACTGATGTGATAAGTCAATGAATTTGAAGTTGGATGATGAGGCA----- 1188

QY 320 LeuGluLysLysMetArgLysSerPheThrLeuAspAspPheLeuGluGlnLeuAsp 339
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1189 CTTATAGAGAAGTTGAAGCAGCGTCACTTTTACATTCGAGACATGATGATGACAGTTTCAG 1248

QY 340 GlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsn--- 358
: : : : : : : : : : : ||| : : : : : : : : : : : : : : : : : : : : :
Db 1249 AATATTATGAATGGGCCCAATTCAGTCAGATATTGGGATGATCTCTGGCTTTGGCACA 1308

QY 359 -----LysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIle 376
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1309 GATTTTATGAGCAAGGGAATGACGAGGATCAATGGCA-----AGGCTGAAGAACTG 1362

QY 377 LysAlaIleIleGlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnVal 396
||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 1363 ATGACAATCATGCAGATGATGAACGATCAAGAACTGGACAGTACAGATGCTGCCAAGGTT 1422

QY 397 -----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1423 TTCAGTAAGCAACCAAGGAGATCCAAAGAGTTGCCGGGGATCAGGTGTCACAAGA 1482

QY 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGln----- 429
: : : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1483 GATGTTCAAGAACTCTTGACCCAGTATATACCAAGTTTGGCACAGATGGTCAAAAAGATGGGA 1542

QY 430 -----PheThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 444
||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||
Db 1543 GGTATCAAAAGGACTTTTCAAAAGGCGGTGATGTCCTAAGAATGTGAGTCATGCACAGATG 1602

QY 445 -----GlnAsnMetLeuLysGlyMetAsn 452
||| ||| ||| ||| ||| ||| : : : : :
Db 1603 GCAAAATTAACCAACCAATGCGCCAAATGATGGAC 1638

RESULT 7
BH394904
LOCUS BH394904
DEFINITION AG-ND-138B11.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138B11
, DNA sequence.
ACCESSION BH394904
VERSION BH394904.1 GI:17341045
KEYWORDS GSS.
SOURCE African malaria mosquito.

```

```

ORGANISM Anopheles gambiae
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
          Anopheles.
REFERENCE 1 (bases 1 to 688)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-138B11.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
          Location/Qualifiers
          source          1..688
                        /organism="Anopheles gambiae"
                        /strain="PEST"
                        /db_xref="taxon:7165"
                        /clone="AG-ND-138B11"
                        /note="vector: pECBAC1; Site_1: HindIII"
BASE COUNT 223 a 118 c 154 g 193 t
ORIGIN
Alignment Scores:
Pred. No.: 1.29e-55 Length: 688
Score: 594.50 Matches: 129
Percent Similarity: 73.39% Conservative: 42
Best Local Similarity: 55.36% Mismatches: 55
Query Match: 26.14% Indels: 7
DB: 17 Gaps: 1
US-09-943-108A-2 (1-455) x BH394904 (1-688)
QY 60 AlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleVal 79
||||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 2 GCCTTAGCACAAAATGTTTAAACATCCATTAACCCCGGACAAATTCATCACAATAATGTT 61

QY 80 GlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLys 99
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 CACGATGAGTGTAGTGAATTAATGAGGAGAACCAATGAAGGCTTAATCTTCCGAAAAA 121

QY 100 ProProThrValValMetMetValGlyLeuGluGlyAlaGlyLysThrThrAlaGly 119
||||| : : : : : : : : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ---CCACAAATTAATGATTCGAGGTTTACAGGTTTCTGGTAAACACATCTCCCGGG 178

QY 120 LysLeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAsp 139
||||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 AAATTAGCAAACTCTCTGAAGAGAAAAGAGCAAAATCCGTTATTGGTAGCATGTGAC 238

QY 140 IleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIlePro 159
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GTTTATGCTCTCTCGGATTCATCAGCTAAAGATATTAGGTAGTCAGACAGGGGTCTCT 298

QY 160 ValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHis 179
||||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GTATATACTGAGGAGGTAAACCTTAATCCGTTTCAGATTCTCAGAATGCAATAGAAATT 358

QY 180 AlalysGluGluHisLeuAspPheValIleIleAspThrAla-GlyArgLeuHisIleAs 199

```

```

|||||:  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 GCAAGCAAAATAGCACGATGTTATCATATACCGCAAGGCGTTGGCTATCGA 418

QY 199 pCUAlaLeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLleu 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 TGAAGATGATGAACGAATTTAAACGTTCCACCGTCTGTAACACCTACAGAACTCT 478

QY 219 eLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 TTTTCGTAAGTACTATGACTGCGGAGGATGCTGTGAATCAGCAAAAGCTTTCAAG 538

QY 239 sGlnLeuAsp-ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg-GlyGI 238.
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 ATGTCCTTAATATACGGTGTAGTCTTACTAAATTAGATGCTGATCTCGGAGGTGG 598

QY 258 yAlaLeuLeuSerLeuArgSer-ValThrGlnLysProLysPheVal-GlyMetSer 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 GCCCGCATACCAATCCGTCGGGTAGTAAACCAATTAAGTTATTTTCACCGGT 658

QY 278 GluLysLeuAspGlyLeuGluLeuPhe 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 GAGAAATGGATGCTCTGGGATATTT 685

RESULT 8
BE432304
LOCUS BE432304
DEFINITION EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
ACCESSION BE432304
VERSION BE432304.1 GI:9430147
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 605)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source 1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLE37M2"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCUadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 190 a 97 c 159 g 159 t
ORIGIN

Alignment Scores:
Pred. No.: 3 86e-52 Length: 605
Score: 562.50 Matches: 105
Percent Similarity: 75.25% Conservative: 47
Best Local Similarity: 51.98% Mismatches: 49

```

```

Query Match: 24.74% Indels: 1
DB: 10 Gaps: 1
US-09-943-108a-2 (1-455) x BE432304 (1-605)

QY 65 ValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGluLeuThr 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 TTGATTCGAGGTGTAACACCAAGTAACTAGTTAAGACTGTGCTGACGAGCTTG 61

QY 85 LysLeuMetGlyGlyGlnThrSerIleAsnMetSerAsnLysProThrValVal 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 AAATCTGATGGTGGAGAGTTCTTGAATGTTTTCGTAATCTAAGCCACCTAATA 121

QY 105 MetMetValGlyLeuGlnGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 CTAATGCCGGTCTACAAAGTGTGGGAGCAACTTTAGTCGCAAGTTAGCTTATAT 181

QY 125 MetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleThrArgProAla 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 CTA---AAGAAGCAGGGTAAGAGTTGCTGCTGATTCCTGGAGACGTGTACAGACCTGCT 238

QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGly 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 GCTATGTACCAACTTGTATTGTTGGTAAACAGGTTGATCTCTTTTATGACGACGA 298

QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 ACAGATGTAAACCTGCAGAAATAGCCCGCAAGGATTACAGAGCCCAAAAGAGAT 358

QY 185 LeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 GTAGATGTAGTCAATTAATGGATACAGCTGGAAGACTTCAGATAGATAAACTATGATGAT 418

QY 205 GluLeuLysGlnValLysGluLeuAlaLysProAsnGluLeuMetLeuValValAspSer 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 GAATTAAGACCTGGAACGGGTACTGAACCCACAGAGGTTTACTTCTTGTGATGCA 478

QY 225 MetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThr 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 ATGACTGCCCAAGAGCTGCAGCTTGTGTCACACATCTCAATCTCGAAATTTGGAATTA 538

QY 245 GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerLeuArg 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 GTGTCATCTTGACGAAGTAGATGGAGATTCTAGGGGTGGAGCAGCTTTAAGTGTCA 598

QY 265 SerVal 266
|||||
Db 599 GAGGTA 604

RESULT 9
AW648731
LOCUS AW648731
DEFINITION EST327101 tomato germinating seedlings, TAMU Lycopersicon
esculentum cDNA clone cLEI522 5', mRNA sequence.
ACCESSION AW648731
VERSION AW648731.1 GI:7409885
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 593)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Niernan,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University

```

100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES

source
Location/Qualifiers

1. .593
/organism="Lycopersicon esculentum"
/cultivar="PA496"
/db_xref="taxon:4081"
/clone="GLE15G22"
/clone_lib="tomato germinating seedlings, TAMU"
/tissue_type="whole seedlings"
/dev_stage="7 days post imbibition"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; 7 days post imbibition on water-agar. Mixed stage whole germinating seedlings from seed coat emergence up to two centimeters in seeds not showing obvious signs of germination were discarded."
BASE COUNT 191 a 97 c 159 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 1.19e-51 Length: 593
Score: 558.00 Matches: 104
Percent Similarity: 76.14% Conservative: 46
Best Local Similarity: 52.79% Mismatches: 47
Query Match: 24.54% Indels: 0
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x AW648731 (1-593)

Qy 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnVal 167
Db 2 CRACCTGTTATTGGGTAAACAGCTTGATGACTGCTGTTATGACGAGCAAGATGTA 61
Qy 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPhe 187
Db 62 AAACCTGCAGAAATAGCCGACAGGATTACAGAGCCCAAAAGAAAGATGTAGATGA 121
Qy 188 ValIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLys 207
Db 122 GTCATATGATGATGACCTGGAGACTTCAGATAGATAAACTATGATGATGATGATGAT 181
Qy 208 GluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMetThrGly 227
Db 182 GACGTGAACGGGTACTGAACCCACAGAGCTTTACTTGTGTGGATGCAATGACTGCC 241
Qy 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247
Db 242 CAAGAAGCTGCAGCTTTGGTCACAACTTCATCTCGAAATGGAATTAAGTGGCCATC 301
Qy 248 LeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSerIleArgSerValThr 267
Db 302 TTGACGAAGCTAGATGGAGATCTAGGGGTGGAGACCTTAAGTGTCAAGGAGGATCA 361
Qy 268 GlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHis 287
Db 362 AGAAGCCATCACTCGTAGGAGGGGTGACCTATGAGGACCTTGAACCTTCTAT 421
Qy 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLys 307
Db 422 CCTGACCGCATGGCTGGACGTATTTAGGAATGGGAGATGTTCTATGCTTTGTTGAGAA 481
Qy 308 AlaGlnGluAspValAspGlnGluLysAlaLysAspLeuGluLysMetArgGluSer 327
Db 482 GCCCAAGAGTATTCGTCAGAGAAGATGCTGAAGATCTCGAAGAAGATCAATGATGCA 541
Qy 328 SerPheThrLeuAspPheLeuGluGlnLeuAspGlnValLysAsnLeu 344
Db 542 AAATTGATGCAATGACTTCTCGAAGCAAACTCCGGCAGCTGCTAGGATG 592
RESULT 10
AI977909
LOCUS AI977909 581 bp mRNA linear EST 27-AUG-1999

DEFINITION

496035E06.x1 496 - stressed shoot cDNA library from Wang/Bohnert lab Zea mays cDNA, mRNA sequence.

ACCESSION AI977909
VERSION AI977909.1 GI:5791117
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 581)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 496035 row: E column: 06.
Location/Qualifiers
1. .581
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="496 - stressed shoot cDNA library from Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E.coli XL Gold"
/note="Organ: shoot; Vector: pBluescriptII SK(+) XR;
Wang/Bohnert"

FEATURES

source
1. .581

/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="496 - stressed shoot cDNA library from Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E.coli XL Gold"
/note="Organ: shoot; Vector: pBluescriptII SK(+) XR;
Wang/Bohnert"

BASE COUNT 176 a 93 c 161 g 150 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2.19e-51 Length: 581
Score: 555.50 Matches: 105
Percent Similarity: 74.61% Conservative: 39
Best Local Similarity: 54.40% Mismatches: 48
Query Match: 24.43% Indels: 1
DB: 9 Gaps: 1

US-09-943-108a-2 (1-455) x AI977909 (1-581)

Qy 38 AlaLeuPheGluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSer 57
Db 4 GCACCTTTGGAGCGCATGTAAAGTTGCCAGTAGTAAGAAGATTTATTGAGTCTGTAAGT 63
Qy 58 GluArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLys 77
Db 64 GAAAGGCTGTAGGACCGCATGTATCCGAGGTGCCACCTGCACGAGTGGTGAAG 123
Qy 78 IleValGluAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSer 97
Db 124 GTTGTGAGTGTGACTGGTACAACTATGCGCGGGAGGATCATCAGATTTGGTGTGCA 183
Qy 98 AsnLysProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrThr 117
Db 184 AAATCTGGCCCAACTATTATTATTTGGCAGGCTGCAAGGTGTTGGAATAACTACTGT 243
Qy 118 AlaGlyLysLeuAlaLeuMetArgLysLysTyrAsnLysLysProMetLeuValAla 137
Db 244 TGTGCGAAGCTTGCTTCTATCTC---AAAAAAGTGGCAAGAGCTGTATGCTAGTGT 300
Qy 138 AlaAspIleThrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAsp 157
Db 301 GCAGATGTTTACAGGCTGCTGCTATTTGATCAACTCATCTTCTGGGTAAAGGTTGT 360
Qy 158 IleProValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeu 177

```

102 361 GTACCAAGTTTACTCAGAGGAACCTTGAAGCAAAACCTTCACAAATAGCCAAAACGGTTTG 420
103 178 LysHisAlaLysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHis 197
104 421 AAGAGGCAAAAGTCAATAGGCTGATGATATATAGTGGACAGCGCTGGAAGACTGCAG 480
105 198 IleAspGluAlaLeuMetAsnGluLysGluValLysGluIleAlaLysProAsnGlu 217
106 481 GTAGTAAGCAATCATGATGATGAGTGAAGAGAGTAAAGAAAGACAGTGAATCCTACAGAA 540
107 218 IleValLeuValLysSerMetThrGlyGlnAspAla 230
108 541 GTTCTCTTGGTGTGATGCCATGACTGGCCAGAGACT 579
109
110 RESULT 11
111
112 LOCUS BM413418 677 bp mRNA linear EST 22-JAN-2002
113 DEFINITION EST587745 tomato breaker fruit Lycopersicon esculentum cDNA clone
114 CLE6363C14 5' end, mRNA sequence.
115
116 ACCESSION BM413418
117 VERSION BM413418.1 GI:18265048
118 KEYWORDS EST.
119 SOURCE tomato.
120 ORGANISM
121 Lycopersicon esculentum
122 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
123 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
124 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
125 Lycopersicon.
126 REFERENCE 1 (bases 1 to 677)
127 AUTHORS Alcala,J.J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
128 J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Renning
129 C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
130 Generation of ESTs from tomato fruit tissue, breaker stage (2002)
131 JOURNAL Unpublished (2002)
132 COMMENT Contact: CUGI
133 Clemson University Genomics Institute
134 100 Jordan Hall, Clemson, SC 29634, USA
135 Email: http://www.genome.clemson.edu/orders/index.html
136 This clone is available through the Clemson University Genomics
137 Institute
138 Seq primer: T3.
139
140 FEATURES
141 source Location/Qualifiers
142 1..677
143 /organism="Lycopersicon esculentum"
144 /cultivar="TA496"
145 /db_xref="taxon:4081"
146 /clone="CLE63C14"
147 /tissue_type="Pericarp"
148 /dev_stage="breaker"
149 /lab_host="SOLR"
150 /note="Vector: pBluescriptSKmCquadapt; Site.1: EcoRI;
151 Site.2: XhoI; supplier: Boyce Thompson Institute; Fruit
152 sequencing: The Institute for Genomic Research. Fruit
153 were harvested at the breaker stage (first sign of
154 lycopene accumulation on the blossom end of fruit). Fruit
155 were cut in half and the seeds and locules were discarded
156 prior to freezing the pericarp."
157
158 BASE COUNT 222 a 119 c 165 g 171 t
159 ORIGIN
160
161 Alignment Scores:
162 Pred. No.: 6 42e-49 Length: 677
163 Score: 534.00 Matches: 101
164 Percent Similarity: 75.38% Conservative: 46
165 Best Local Similarity: 51.79% Mismatches: 47
166 Query Match: 23.48% Indels: 1
167 DB: 13 Gaps: 0
168
169 US-09-943-108a-2 (1-455) x BM413418 (1-677)

```

```

170 152 ValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnValLysProGlnGln 171
171 92 GTTGGGGAACAGTTGTGATCTTATGCGAGGACAGATGATAAACCTCGAGAA 151
172 172 IleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPheValIleLeuAsp 191
173 152 ATAGCCCGACAAAGGATTACAGAGGCGCAAAAGAAATAGTAGTGTAGTCAATAATGGAT 211
174 192 ThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLysGluValLysGlu 211
175 212 ACAGCTGGAGAGCTTCAGATAGATANAACATGATGATGATTAAGAGACGTGAACGG 271
176 212 IleAlaLysProAsnGluIleMetLeuValLysSerMetThrGlyGlnAspAlaVal 231
177 272 GTACTCAACCCACAGAGGTTTACTTGTGTGGATGCAATGACTGGCCAGAGAGTGCA 331
178 232 AsnValAlaGluSerPheAspAspGlnIleLeuAspValThrGlyValThrLysLeu 251
179 332 GCTTGGTCCACAACTTCATCTCGAAATTTGGAATTTACTGTGCGCATCTTGACGAAGCTA 391
180 252 AspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIle 271
181 392 GATGGAGATCTAGGGGTGGAGCAGCTTAAGTCTCAGGAGGTATCAGGAAGCCCAATC 451
182 272 LysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMet 291
183 452 AAGCTCGTAAGAGGGGTGAACGTATGGAGGACCTTGAACCTTTCTATCTGACCGCATG 511
184 292 AlaSerArgIleLeuGly-MetGlyAspValLeuSerLeuIleGlyLysAlaGlnGln 311
185 512 GCTGGAGCTATTATTAGGAATTTAGGAGATGTTCTATCGTTGTTGAGAAAGCCCAAGAG 571
186 311 pValAsnGlnLysAlaLysAspLeuGluLysLysMetArgGluSerPheThrLe 331
187 572 TATCGCTCAGAGAGATGCTCGAAGATTTCAGAGAAGATCATGAGTGCATAAATTGATT 631
188 331 uAspAspPheLeuGluGlnIleLeuAspGlnValLysAsnLeuGly 345
189 632 CAATGACTTCTCTGAAGCAAACTCCGCGAGTTGCTAGTAGTGGST 674
190
191 RESULT 12
192
193 LOCUS AY109321 1200 bp mRNA linear HTC 25-MAY-2002
194 DEFINITION Zea mays PC0070230 mRNA sequence.
195 ACCESSION AY109321
196 VERSION AY109321.1 GI:21212876
197 KEYWORDS HTC.
198 SOURCE Zea mays.
199 ORGANISM
200 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
201 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
202 clade; Panicoideae; Andropogoneae; Zea.
203 REFERENCE 1 (bases 1 to 1200)
204 AUTHORS Hailey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
205 Arthur,J.W., Hanafey,M., Morgante,M. and Tingey,S.V.
206 Maize Mapping Project/DuPont Consensus Sequences for Design of
207 Overgo Probes
208 JOURNAL Unpublished (2002)
209 REFERENCE 2 (bases 1 to 1200)
210 AUTHORS Coe,E.C.
211 TITLE Direct Submission
212 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
213 Missouri, Columbia, MO 65211, USA
214 FEATURES
215 Location/Qualifiers
216 1..1200
217 /organism="Zea mays"
218 /db_xref="taxon:633818"
219 /db_xref="maizeDB:633818"
220 /clone_lib="pcc070230"
221 /clone_lib="Maize Mapping Project/DuPont Consensus
222 Library"
223 /note="this sequence is part of a project of EST
224 assemblies resulting from the application of public

```

contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"

BASE COUNT 387 a 216 c 317 g 279 t 1 others
ORIGIN

Alignment Scores: 4.15e-47 Length: 1200
Pred. No.: 520.50 Matches: 102
Score: 70.82% Conservative: 63
Percent Similarity: 43.78% Mismatches: 65
Best Local Similarity: 22.89% Indels: 3
Query Match: 11 Gaps: 2
DB:

US-09-943-108a-2 (1-455) x AY109331 (1-1200)

QY 208 GlnVallysluileAlaLysProAsnGluileMetLeuValValAspSerMetThrGly 227
DB 11 GAAGTGAAGAAGCTGTTAACTCCACAGAAGTTCGTGCTGATGCGCATGCTGGC 70
QY 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247
DB 71 CAAGAAGCAGACACACTAGTCACCACTTCAATATTGAGATTGGTATTAACCTGGTGAATA 130
QY 248 LeuThrLysLeuAspGlyAspThrArgGlyAlaAlaLeuSerIleArgSerValThr 267
DB 131 CTGACTAAATGGATGGTGACTCCAGGGCGGAGCGGAGCTAAGCGTTAAAGAGGCTCT 190
QY 268 GlnLysProIleLysPheValClyMetSerGluLysLeuAspGlyLeuGluLeuPheHis 287
DB 191 GGGAGGCCATAAAATTTGTTGGCGTGGGACGATAGAGGACCTTGACCTTTCTAC 250
QY 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuLleGluLys 307
DB 251 CCGTATCGATGCCACAGCAGAGTCTCTGGGAATGGGAGAAGTGTCTCATTTGTGCAAAAG 310
QY 308 AlaGlnGlnAspValAspGlnLysAlaLysAspLeuLysLysMetArgGluSer 327
DB 311 ACACAGCAATGATGAGGCGAGAGAGCTATAGAAATTACAGAAGATCACTGAGTGCA 370
QY 328 SerPheThrLeuAspAspPheLeuGlnLeuAspGlnValLysAsnLeuGlyProLeu 347
DB 371 AAATTCGACTTCACAGATTCTTAAACAATCTCAAAACGTTGCGAAAAATGGGTTCATG 430
QY 348 AspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsn 367
DB 431 AGCCGCATTATTGGAATGATGCCAGGCTGAACAAGATCACTCCCGCA---CAAATCCGA 487
QY 368 MetSerGluLysGlnLysPheHisIleLysAlaIleIleGlnSerMetThrProAlaGlu 387
DB 488 GAACCTGAGAAAGACATTCATCTGGAGTCATATGATCAATGCCATGCTGCTGAGGAA 547
QY 388 ArgAsnAsnProAspThrLeuAsnValSerArgLysLysArgIle-----AlaLysGly 405
DB 548 AGGAGAGACCGACAGTACTGGCTGAATCACGTGAGAGGAGGATAAGAGTGGCTGAGGAG 607
QY 406 SerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
DB 608 TCTGGAAGACTGAACAAGAGGTGATCAATTGGTTGCCACGCTTTTCCAAATCGGTCT 667
QY 426 MetLysGlnPheThrGlyGlyGlyLysLysLys 438
DB 668 CAGATGCAAGAGTGTATGGGTATGTCGCAAGGACAGAA 706

RESULT 13
AW930742
LOCUS AW930742 533 bp mRNA linear EST 18-MAY-2001
DEFINITION EST356585 tomato fruit mature green, TAMU Lycopersicon esculentum
cDNA clone cLEF40N15 5', mRNA sequence.
ACCESSION AW930742
VERSION AW930742.1 GI:8106143
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 533)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Renning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1..533
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF40N15"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"
BASE COUNT 171 a 86 c 144 g 131 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 8,86e-47 Length: 533
Score: 513.50 Matches: 97
Percent Similarity: 76.27% Conservative: 38
Best Local Similarity: 54.80% Mismatches: 41
Query Match: 22.58% Indels: 1
DB: 10 Gaps: 1

US-09-943-108a-2 (1-455) x AW930742 (1-533)

QY 108 GlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuMetArgLys 127
DB 5 GGTCTACAGGTGTGGGAAGACAACTGTTAGTGCAAAGTAGCTTATATCTA--AAG 61
QY 128 LysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaLeuAsn 147
DB 62 AAGCAGGTAAGAGTGGATGCTGATTCGTGGAGACGTGTACAGACCTGCTGCTATTGAC 121
QY 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluLysAspGlnVal 167
DB 122 CAACTGTTATTTGGTAAACAGGTGTACCTGTTTATGTCAGCAGGACAGATGTA 181
QY 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAspPhe 187
DB 182 AAACCTGCAGAAATAGCCGCAAGAGATTACAAGAGGCCAAAGAGAGATGATAGTGA 241
QY 188 ValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLys 207
DB 242 GTCATAATGGATACAGCTGGAGACTTCAGATAGATAAAACATATGATGATGAATTA 301
QY 208 GlnValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMetThrGly 227
DB 302 GACGTGAACCGGTACTGAACCCACAGAGGTTTACTTGTGTCGATGCAATGACCTGGC 361
QY 228 GlnAspAlaValAsnValAlaGluSerPheAspGlnLeuAspValThrGlyValThr 247

```

Db 362 CAAGAAGCTCCAGCTTTGGTCACAAATCAATCTCGAAATTGGAAATTACTGGTCCATC 421
QY 248 LeuThrLysLeuAspGlyAspThrArgGlyValAlaLeuSerLeuArgSerValThr 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 422 TTGACGAGCTAGATCGAGATCTCAGGGGUGGAGCAGCTTTAAGTGTCAAGAGGTATCA 481
QY 268 GlnLysProTleLysPheValMetSerGlnLysLeuAspGlyLeuGlu 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 482 NGAAGCAATCAAGCTCGTAGGAGGGGTGAACGTATGAGGACCTTGAA 532

RESULT 14
BQ996389 677 bp mRNA linear EST 22-AUG-2002
LOCUS QGG12L07.Yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
DEFINITION QGG12L07.Yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
ACCESSION BQ996389
VERSION BQ996389.1 GI:22430785
SOURCE EST.
ORGANISM Lactuca sativa.
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 677)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmil.ucdavis.edu]
belongs to contig QG_CA_Contig5305, see http://cgpdb.ucdavis.edu/
for details.
Plate: QGG12 row: L column: 07.
FEATURES
source
location/Qualifiers
1..677
/organism="Lactuca sativa"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGG12L07"
/lab_host="E.Coli"
/note="Vector: pBRCNDSFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=CGTGCAGGG"
BASE COUNT 213 a 100 c 192 g 172 t
ORIGIN

Alignment Scores:
Pred. No.: 7.17e-46 Length: 677
Score: 506.50 Matches: 100
Percent Similarity: 71.43% Conservatives: 45
Best Local Similarity: 49.26% Mismatches: 57
Query Match: 22.27% Indels: 1
DB: 14 Gaps: 1

```

```

US-09-943-108A-2 (1-455) x BQ996389 (1-677)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetClnLysMetArgGlyLysGlyLys 22
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 72 TTTGGTCAGCTAACCCAGTGGCCCTTGAATCCGCCCTGGGAACAAGCTTAAGGGCGAAGAGGTC 131
QY 23 LeuThrGluAlaAspLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 132 TTAAGGAGGAGGATATCTCGAACCAATGAGGACATTAGACGAGCTCTTCTGGAGCA 191
QY 43 AspValAsnPheLysValLysGluPheLleLysThrValSerGluArgAlaLeuGly 62
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 192 GATGTAGCTCTCTCTGTTAAGAAGATCTGTTCAAACTGTAAGTGAACAAGCTGTGGT 251
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 252 GTTGGTGTGACTAGAGGAGTCAACCAAGATCAGCAATGTGTAATCTAGTGTGAA 311
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThr 102
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 312 CTCGTGAAGCTAATGGAGAGAGAGGTTCCTGAGTTAACTTTGCAAAATCTGGCCCACT 371
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 372 GTTATCTTACTGGCTGGTCTACAGGTTGTTGGGAAGACAACCTGTTAGTGCAAAATTAGCT 431
QY 123 LeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspLleTyArg 142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 432 TTGTACCTC---AAGAAACAGGGGAGAGTTCGATGCTGATGCTGGAGATGTAATAGA 488
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspLleProValTySer 162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 489 CCAGCTGCGATTGATCAACTGTTATTTGGTAAACAGGTGATGCTTCTGCTGTATGCA 548
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 549 GTAGGAGCTGATGTAAACCAAGCAAGTACCTAGACAAGGCTTCAAGAGGCTAAAAAG 608
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 609 AAGATGTAGATGTGTTATATGATGATACACAGGAGAGCTTCAGATAGACAAAATATG 668
QY 203 MetAsnGlu 205
||||| |||||||
Db 669 ATGGATGAA 677

RESULT 15
BQ323434
LOCUS BM323434
DEFINITION PIC1_19_C04.b1_A002 pathogen-infected compatible 1 (PIC1) Sorghum
bicolor cDNA, mRNA sequence.
ACCESSION BM323434
VERSION BM323434.1 GI:18061215
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 514)
Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
Sudman,M. and Pratt,L.H.
An EST database from Sorghum: plants infected with a compatible
pathogen
JOURNAL Unpublished (2002)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector, and regions

```

1.                                                            

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 00:20:56 ; Search time 3139 Seconds
(without alignments)
4218.470 Million cell updates/sec

Title: US-09-943-108a-2

Perfect score: 2274

Sequence: 1 MAFGLSERLQATQAKMKK.....GKKGRNQMNLMKGNLPF 455

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-O=/cgn2_1/USPTO.spool/US0943108/runat_21022003_151630_1801/app_query.fasta_1.647
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0943108.cgn_1_1.3745,runat_21022003_151630_1801 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rnd.*
36: em.htg_nam.*
37: em.htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2274	100.0	1368	6	E36050	E36050 Signal reco
2	2274	100.0	303750	1	AP003133	AP003133 Staphyloc
3	2274	100.0	347235	1	AP003361	AP003361 Staphyloc
4	2261	99.4	295350	1	AP004826	AP004826 Staphyloc
C 5	1719.5	75.6	2256	6	AX416333	AX416333 Sequence
C 6	1717.5	75.5	347050	1	AL591981	AL591981 Listeria
C 7	1708.5	75.1	313450	1	AL596170	AL596170 Listeria
8	1708.5	75.1	319630	6	AX413016	AX413016 Sequence
C 9	1708.5	75.1	34980	6	AX417046	AX417046 Sequence
10	1647	72.4	2739	1	D14356	D14356 Bacillus su
11	1647	72.4	208780	1	BS080009	Z99112 Bacillus su
C 12	1584	69.7	303249	1	AP001515	AP001515 Bacillus
13	1402	61.7	1136	6	AX433432	AX433432 Sequence
14	1393	61.3	11024	1	AB007684	AB007684 Clostridi
C 15	1382	60.8	296750	1	AP003191	AP003191 Clostridi
C 16	1352.5	59.5	14571	1	AE013104	AE013104 Thermoana
C 17	1315	57.8	11023	1	AE006391	AE006391 Lactococc
C 18	1312	57.7	1605	6	AX194269	AX194269 Sequence
C 19	1307	57.5	10462	1	AE008489	AE008489 Streptoco
C 20	1306	57.4	1572	6	AX194043	AX194043 Sequence
C 21	1306	57.4	7577	6	BD003726	BD003726 Polyucle
C 22	1306	57.4	9985	1	AE007428	AE007428 Streptoco
23	1305	57.4	792	6	E36051	E36051 Signal reco
C 24	1299	57.1	77743	2	SPNEU1910	AL49932 Streptoco
25	1297	57.0	1569	6	AR081886	AR081886 Sequence
26	1297	57.0	1569	6	AR195303	AR195303 Sequence
27	1297	57.0	1569	6	B35383	B35383 Novel fff.
28	1286	56.6	4152	1	U88582	U88582 Streptococc
C 29	1285	56.5	50946	1	AE044153	AE044153 Streptoco
C 30	1283	56.4	10173	1	AE010040	AE010040 Streptoco
C 31	1279.5	56.3	12187	1	AE006560	AE006560 Streptoco
C 32	1238	54.4	11825	1	AE010644	AE010644 Fusobacte
C 33	1180.5	51.9	15721	1	AE001802	AE001802 Thermotog
C 34	1173.5	51.6	21387	1	AE008821	AE008821 Salmonell
C 35	1173.5	51.6	274050	1	AL627276	AL627276 Sequence
36	1168	51.4	1641	6	AX122339	AX122339 Sequence
C 37	1168	51.4	337200	1	AP005280	AP005280 Corynebac
C 38	1168	51.4	349980	6	AX127150	AX127150 Sequence
39	1164.5	51.2	4586	1	ECTRMD	X01818 E. coli trm
C 40	1164.5	51.2	10267	1	AE005491	AE005491 Escherich
C 41	1164.5	51.2	11272	1	AE000347	AE000347 Escherich
C 42	1164.5	51.2	270365	1	AP002562	AP002562 Escherich
43	1150.5	50.6	10566	1	AE004142	AE004142 Vibrio ch
C 44	1150	50.6	11506	1	AE004793	AE004793 Pseudomon
45	1139.5	50.1	11004	1	AE013692	AE013692 Versinia

ALIGNMENTS

RESULT 1

```

E36050      1368 bp      DNA      linear      PAT 18-JUN-2001
DEFINITION  Signal recognition particle polypeptide and polynucleotide.
ACCESSION   E36050
VERSION     E36050.1  GI:13022452
KEYWORDS    JP 1999235183-A/1.
SOURCE      unidentified.
ORGANISM    unclassified.
REFERENCE   1 (bases 1 to 1368)
AUTHORS    Michael T.B.
TITLE       Signal recognition particle polypeptide and polynucleotide
JOURNAL     Patent: JP 1999235183-A 1 31-AUG-1999;
            SMITHKLINE BEECHAM CORP
COMMENT     OS Unidentified
            PN JP 1999235183-A/1
            PD 31-AUG-1999
            PF 03-SEP-1998 JP 1998289963
            PR 03-SEP-1997 US 60/057890, 05-MAR-1998 US 09/035382 PI
            MICHAEL TERENCE BLACK
            PC C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,
            A61K31/00,
            PC A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K35/76,
            A61K39/085,
            PC A61K39/085,A61K39/395,A61K48/00,C07K14/31,C07K16/12,C12N1/15,
            C12N1/19,
            PC C12N1/21,C12N5/10,C12N5/10,C12P21/02,C12Q1/68,G01N33/15, PC
            G01N33/566//
            PC C12P21/08,G01N33/577,(C12N15/09,C12R1:445),C12N15/00,C12N5/00,
            C12N5/00,C12R1:445)
            CC Strandedness: Single;
            CC Topology: Linear;
            FH Key
            FT source 1..1368 Location/Qualifiers
            FT /organism='Unidentified'.

FEATURES             source
     source           1..1368
                        /organism="unidentified"
                        /db_xref="taxon:32644"
BASE COUNT  521 a 186 c 288 g 373 t
ORIGIN
Alignment Scores:
Pred. No.:      6.44e-130      Length:      1368
Score:          2274.00      Matches:      455
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels: 0
DB:             6      Gaps: 0

US-09-943-108a-2 (1-455) x E36050 (1-1368)

QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTGGAAGCTTTATCAGACGCTTCGCAAGCGACGATGCAAAAATGCGTGGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspPileLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAAACCTTACTGAAGCTGTATTAAGATATATGATGCGTGAAGTAGATTACGTTATTT 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPhelLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACGCGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
Db 181 TTAGGTTCCGATGTAATGCAATCATTACACCGAGGCAACAAGTATTATTAATAGTTCAA 240
QY 81 AspGluLeuThrLysIleuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTACGAAGAGTTGATGGGTGGAGAAATATACATCGATTATATATGTCAAATACCA 300

```

```

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGGTTGGTTTACAAGTGCTGCTGTAACAAACAACAACTGCAGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspPile 140
Db 361 TTAGCATTTATTTGATGCGTAAAAAATACAAACAAAACCTATGTTAGTTGCACAGATAAT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspPileProVal 160
Db 421 TATCGTCCAGCAGCGATAAATCAATTACAACAGTACGGAAACAATAATTGATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCACACAAATTTGTAACATAATGATTAATAAATCATCT 540
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTTGTATTCATTGATACAGCAGGTCGATTACACATCGATGAA 600
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleXMetLeu 220
Db 601 GCATTGATGAACGAATTTAAAGAGATTAAGAAATTCGTTAAACCAACGAATTTATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGGAATTCATGTACGGGTCAAGATGCTGCAATGTTGCAGAAATCTTTTGCACATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGTGATGCACAGGTGTTACCTTAACTAATAGATGTTGATACAGTGTGTGTCGTCACCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCTATTCGTCGTCGACACAAAAACAATTAATTTCTGCTATGAGTCAAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGTTTTAGAGCTTATTCATCTCTGACGATGTCATCAGTATTTAGGTATGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAAATTGAAAAGCGCCACACAGATGIGGATCAAGAAAACCAAAAGATTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAGATGCGTGAGTCAATCGTTTACTTTAGATGATTTTGTAGAACACACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCCACTGGATGATATTTATGAAAATGATTCACAGGTATGAATAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTAGATTAAGCTTAATATGAGTGAAAGCAAAATGATCATATTATAACGATATTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACCGCGCTGAAAGAACAAATCCACACATGGAATGATCATCGTAAATAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGCTCTGTCGTTCAATACAAAGTCAATCGTTTGTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLysLys 440
Db 1261 AACGATATCAAGAAAATATGATGAACCAATTCATGCTGGTGGCGGTAAAGGTAAAAAGGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365
RESULT 2

```

```

AP003133      303750 bp      DNA      linear      BCT 02-JUL-2002
LOCUS      Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
DEFINITION      genome, section 5/10.
ACCESSION      AP003133 BA000018
VERSION      AP003133.2 GI:14349175
KEYWORDS
SOURCE      Staphylococcus aureus subsp. aureus N315 (sub_species:aureus N315,
      strain:N315) DNA.
ORGANISM      Staphylococcus aureus subsp. aureus N315
REFERENCE      Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS      Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
      Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
      Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A., Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
      Sekimizu,K., Hiramatsu,H., Kuhara,S., Goto,S., Yabuzaki,J.,
      Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
      Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
      Whole genome sequencing of methicillin-resistant Staphylococcus
      aureus
Lancet 357 (9264), 1225-1240 (2001)
MEDLINE      21311952
PUBMED      11418146
REFERENCE      2 (bases 1 to 303750)
AUTHORS      Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K.
      and Kikuchi,H.
TITLE      Direct Submission
JOURNAL      Submitted (30-JAN-2001) Akio Oguchi, National Institute of
      Technology and Evaluation, Biotechnology Center; 2Chome 49-10
      Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
      (E-mail:oguchienite.go.jp, URL:http://www.bio.nite.go.jp/,
      Tel:81-3-3481-8423, Fax:81-3-3481-8424)
      On Jun 12, 2001 this sequence version replaced gi:13701012.
FEATURES
      source
      1. 303750
      /organism="Staphylococcus aureus subsp. aureus N315"
      /strain="N315"
      /sub_species="aureus N315"
      /db_xref="taxon:158879"
      202..1155
      /gene="SA1056"
      /gene="SA1056"
      /note="ORFID:SA1056"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="BAB42308.1"
      /db_xref="GI:13701013"
      /translation="MKKTGCLLLIMLVAGCGFGNHLKSSKSESKQFTVKKES
      EEKDPDEKVEETEKKMGIKDAPSLDKPLMTEKSFNSKGIQGWKDYKELMGKY
      ELADRFNKKSGSSIKDVAFFKGGKKIKKRVLETDDVKQVDYVYDPDGKKIGNS
      NPFVYASIMTKYDKGLVYASVEPGSVYIHKDDAIKYDDYSKKLKSQTLKLDHKRP
      VPSVAQIKSGFVPLTSFEMTHGSDTKDEVLPALEYFFSPKNYEDSKNPDPKVLN
      LVHMFDFNASSDFGNAHFVYLSKYIKEYESNTEVASDLSK"
      complement(1318..1521)
      /gene="SA1057"
      complement(1318..1521)
      /gene="SA1057"
      /note="ORFID:SA1057"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="BAB42309.1"
      /db_xref="GI:13701014"
      /translation="MQWNKVIYILLALFLGSGFIHKFYAGKNMQGILHLIFCWTGIPH
      ILAIIASVITVFKPADEQGNVTL"
      1820..2308
      /gene="SA1058"
      1820..2308
      /gene="SA1058"
      /note="ORFID:SA1058"
      /gene="SA1058"
      /note="ORFID:SA1058"

      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="BAB42310.1"
      /db_xref="GI:13701015"
      /translation="MAKKKLVPAASHPIITKKAQAVKTFDDSLKRLQLQLEDTMYAQEA
      AGLCAPQINSLQVAIIDMEMBGLQLVNPKEISQSNETITDLEGSITLPDVGVEVTR
      SKMIVSYDVNGKVELTAHEDVARMILHIIDQMNGIPTTERADRILLTDKEVEAYFI
      ND"
      2301..3236
      /gene="SA1059"
      2301..3236
      /gene="SA1059"
      /note="ORFID:SA1059"
      /codon_start=1
      /transl_table=11
      /product="methionyl-tRNA formyltransferase"
      /protein_id="BAB42311.1"
      /db_xref="GI:13701016"
      /translation="MTKIIIFMGTPDFSTTVLEMLIAEHDVIAVTPQDPVGRKRVMT
      PPVKKVMKYDLPVYQPEKLSSELEBQLLDVLDIVTAARGQLLESILALPKLG
      AINVASLLPKYGGAPIQALIDGEQETGITMYMKKLDAGNIISQAQKIEENDN
      VGTMDKLSVLGADLLKTELPSIIEGTNESVPQDTQATFASIRREDRISNKPGR
      QVFNQIRGLSPWPVAYTWMDDTNLKIYDAELVETNKINEPGTIIETFKKAIIVATNDN
      EAVAIDKQLGAKGRMAAANYLSGAQNTLVGKKLI"
      3155..4540
      /gene="SA1060"
      3155..4540
      /gene="SA1060"
      /note="ORFID:SA1060"
      /product="hypothetical protein, similar to RNA-binding Sun protein"
      /codon_start=1
      /transl_table=11
      /protein_id="BAB42312.1"
      /db_xref="GI:13701017"
      /translation="MSNEKENVSQOLFKAHAKHTSREETYMIENVSFAFDIQLN
      EGAYENLRINEVSENELNMDKALTEIVTGYTKRYTLDYFKPFVKTIKAWVRQ
      LWMVSIQYVYLDKVPNHAIINEAVEIAKERGGVNGVNGVNGILRTVMKRSIDLPNEI
      ADPKRMAIEYSMPKTIIDHWATHYGLBETETILQSLETTSTTVRANLFRASLDDII
      EKLQDEGYDEKDHDPYCLHIGGQPIIHSRSFQDGSFQDSSMFAHIMNVDRHD
      HVLDAQAPGGKACHAEVLMPGQVQDASDIHDKHDLINFNKKLRNLINIKAFQHDA
      TKYDKTYKILVDAPGSGVGNRHKPEIKYQSKQHIIEVSELVQLEILELVNKNVKI
      GGEIITCTCTQLENEENVITFLNKNKNFEFPFHQITGELVKLTQMPQDFNSOG
      FEITKIKRKN"
      4543..5637
      /gene="SA1061"
      4543..5637
      /gene="SA1061"
      /note="ORFID:SA1061"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="BAB42313.1"
      /db_xref="GI:13701018"
      /translation="MITAEKKKKNKFLPNFQKQSIYSILRDEMQNMLVGGQOKPRAK
      QIFELWKRYSDSIDENLNSKDLQLLDNFTVTLTVVVKQSKDGITKELFELQD
      GTIETVLMRDYNGSVCTVQVCGICTGCTGLGKRNLEAGEIVSVLYTQKA
      LDATERVSOIVMIGICEPENYDNMDFLRIVNDNSLNGARHITVSTSGIIPRY
      DEADEDIQINFAVSLHAAKDEVSRMLPINRAYNVEKLEIAIOYQEKTRNRTVEYG
      LFGVNDQLEHARELAHLKGLNCHVNLIPINVHPYERNYKTAANDLFFKEKELKLS
      INATIRBQSGSDIDAACQQRAKERQVETR"
      5644..6387
      /gene="SA1062"
      5644..6387
      /gene="SA1062"
      /note="ORFID:SA1062"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="BAB42314.1"
      /db_xref="GI:13701019"
      /translation="MLEAQFDTGQHRDKNEDAGGYFQNTNQQLLVLCDSGQGHKA
      GEVASKFVTDELKSRFEAENLIEHQENLNNKIKDINFQLYHQENAEYKGGTIT
  
```

gene 6384. .8378 /gene="SA1063"

CDS 6384. .8378 /gene="SA1063"

/note="ORFID:SA1063"

/codon_start=1

/transl_table=11

/product="protein kinase"

/protein_id="BAB42315.1"

/db_xref="GI:13701020"

CVCALVFSEKSVIANVDSRAYVINSRQIEQITSDHSFVNHLVLTGQITPEBAFTHPQ
RNIITKVMGTDKRYSPDLFTKRLNFYDILLNSDGLTDYVKDNEIKRLLVKGGTIEDH
GQLMLALDNHSDNVTFLAALIEGDKV"

/translation="MIGIKINERYKIYDKLGGGMSVLYLAEDTILNLIKVALFIP
PREKETIKRERVSHNSOLSHONIVSMIDVEDDCYLYVMEYIEGPTLSEYTESH
TSLTVDTAINTFNOILDGIKHAHDMRIVHRDIKPNILIDSNKTLKIFDGLAKLSE
GLSDVTNHLVLTQVYFSPQAGKEATDECTDIYSIGIVLEMLVGPFPNGTAYSIA
IKHIOQSPNVTTRKDIQFQSLNSVILRATEKDKANIKTIQMKDDLLSSVHENRA
NEDVYELDKMTIAVLPKKEDLAKHISEKSNQPKRETTQVIVNGPAPHQQFQKEG
TYTEPKPKKSTRKIVLLSIFLLMIALVSVFAMFNGKYEETPDVIGKVKAEQ
IPNKNLKGKISRSYSDKYPENEIITKTPNTGERVERGSDVDVSKGPEKVPNV
IGLPEALOKLSLGLKDYTIIEKVYVNOAPKGYTANGSVTANTEIAIHDSKILYES
LGTKOVYVEDFEHSFKAKKALEEKPKVESKESYSDDDIDEGDVISQSPKESVDES
SISVSVKSKSSSDVKTTSYDVPITGKNDSKQVAVIINDKNDGSGTEKGSFD
IYSDQRIDIPURIEIKBKGTASYIYKVDGKTVAEKEVSYDDI"

/gene=".9481"

/gene="SA1064"

8606. .9481

/gene="SA1064"

/note="ORFID:SA1064"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="BAB42316.1"

/db_xref="GI:13701021"

/translation="MKTRGVKISGVQVDVNGRPNTRKPLFRKKKSPVWGDIY
REFVONIEGTHQVFKKMLRPPVSNIDTLVIVNSAVEPNFSTQLDRFLVIAHS
YQNLARVLTKKDTPKEQFEINELLKIYENIGYETEFIGNDDDRKKIYEAPAGLI
VLSGQGVKSTFNLNHYRPELNTDIDSKSLNKGKHTTRHVELFERQNGYIADTPGF
SALDPDHIDKDEIKDYFLELNRYGCTCFKFNCHIKPEKNCVHOLEIGNIAQFYDH
YLQLFNETSNRKRY"

gene 9482. .10126

CDS 9482. .10126 /gene="cfxE"

9482. .10126 /gene="cfxE"

/note="ORFID:SA1065"

ribulose-5-phosphate 3-epimerase homolog"

/codon_start=1

/transl_table=11

/protein_id="BAB42317.1"

/db_xref="GI:13701022"

/translation="MTKLXPSLLSVDFDLQHELKRLKEAGVDGVHFDVMDGQVFNII
STGLPDLDAVRGWTLPIDVHLMENPEKYIASFAEHGADMISLHVESPTPHRALQM
IKHLKKAGVINGPTLSIQIEFILDIYIVLVNPNFGQGSFIDOCVEKTAGLNA
IKMEROLNFDIEVDGNTDTAKVCVNGEATMLVTGSFFFKQEDYKKVQTALUG"

Alignment Scores:

Pred. No.:	3	07e-127	Length:	303750
Score:	2274.00	Matches:	455	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	1	Gaps:	0	

US-09-943-108a-2 (1-455) x AP003133 (1-303750)

Qy 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||

Db 26306 ATGGCATTGAAGCTTATCAGACGCTTGCAAGCGACGATGCAGAAAAATGGTGGTAAG 26365
|||||

Qy 21 GlyLysLeuThrGluAlaaspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||

Db 26366 GGTAAACTTACTGAAGCTGATATAAGATAATGTGCGTGAAGTAAGATTATTT 26425
|||||

QY 41 GluAlaaspValasnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||

Db 26426 SAGGCTGACGTAAACTTTAAAGTGGTAAAGAAATTTATTAACACATCATCAAGCGCA 26485
|||||

QY 61 LeuGlySeraspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||

Db 26486 TTAGTTCCTCGATGATGAATCATTAACACGCGGCAACAAGTTATTAATAATAGTCAA 26545
|||||

QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||

Db 26546 GATGAATTAACGAAGTTGATGGGTGGAGAAATATCATCGATTAAATATGTCATAAACA 26605
|||||

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||

Db 26606 CCTACTGTTGTTATGATGTTGTTTCAAGGTGCTGGTAAACACAACTTCAGTAA 26665
|||||

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysProMetLeuValAlaAlaaspIle 140
|||||

Db 26666 TTAGCATTTATTGTCGTGTTAAATAATACACAAAAAACCCTATGTTAGTTGCAGCATAT 26725
|||||

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleaspIleProVal 160
|||||

Db 26726 TATCTCCGCGACGATTAATCAATTAACACATGAGGAGAAATGATGATTCCTCTGA 26785
|||||

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||

Db 26786 TACAGTGAAGGAGATCAAGTAAAGCCACACAATTTGTAACATTAATCAATAAACA 26845
|||||

QY 181 LysGluGluHisLeuaspPheValIleIleaspThrAlaGlyArgLeuHisIleaspGlu 200
|||||

Db 26846 AAAGAAGAACATTTAGACTTTTGTATCAATGATACAGCAGGTGCTGATTACACATCA 26905
|||||

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||

Db 26906 GCATTGATGACAGATTAATAAGAGATTAAGAAATGCTTAACACCAACGAAATTAATGTTA 26965
|||||

QY 221 ValValaspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheaspGln 240
|||||

Db 26966 GTTGTCTGATTCAATGACGGGTCAAGATGCTGTCAATGTTGCAGAACTCTTTGACGATCAA 27025
|||||

QY 241 LeuaspValThrGlyValThrLeuThrLysLeuaspGlyaspThrArgGlyGlyAlaAla 260
|||||

Db 27026 CTTGATGTACAGGTGTTACCTTAATAATTAGATGCTGATACAGTGTGTCAGTGTGTCAGCT 27085
|||||

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||

Db 27086 TTATCTATTCGTCGGTGACACAAAACCAATTAATTTGTTGGTATGAGTGAAAGTTA 27145
|||||

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyasp 300
|||||

Db 27146 GATGTTTTAGAGCTATTCCATCCTGACGATGTCATCAGCTATTTTAGTATGGTGCAT 27205
|||||

QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValaspGlnGluLysAlaLysaspLeu 320
|||||

Db 27206 GTCTTAAGTTTTAATTTGAAAAGCGCACACAGATGTTGATCAAGAAAACCAAGATTTA 27265
|||||

QY 321 GluLysLysMetArgGluSerSerPheThrLeuaspPheLeuGluGlnLeuaspGln 340
|||||

Db 27266 GAGAAAAAGATGCGTGAGTCTGCTTTACTTTAGATGATTTTTTAGAACACATTCATCAG 27325
|||||

QY 341 ValLysAsnLeuGlyProLeuaspPheMetLysMetIleProGlyMetAsnLysMet 360
|||||

Db 27326 GTGAAAAATCTAGGACCCACTGGATGATATTATGAATGATTTCCAGGTATGAATAAATG 27385
|||||

QY 361 LysGlyLeuaspLysLeuAsnMetSerGluLysGlnIleaspHisIleLysAlaIleIle 380
|||||

Db 27386 AAAGGCTTAGATAAGCTTAATATGAGTGAAGGCAAAATGATCATATTTAAACGATTTATC 27445
|||||

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProaspThrLeuAsnValSerArgLysLys 400
|||||

Db 27446 CAGTCAATGACGCGGTGAAGAAACAAATCCAGACATTCATGATGATCAGTAAAAAG 27505
|||||

```

QY 401 ArgilealalysGlySerClvArgSerleuGlnGluValasAcqLeuMetLysGlnPhe 420
|||||
Db 27506 CGTATTCTAAGAGGCTGGTGGTTCATTACAAAGAGTCAATCGTTTATGAACAATTT 27565
|||||
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||
Db 27566 AACCATATGAAGAATAATGATGAACAATTACTGTTGGCGGTAAAGGTAAAGAGGTAA 27625
|||||
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||
Db 27636 CGCAATCAATGCAAAATATGTTTAAAGGTATGAATTTACCGTTT 27670
|||||

RESULT 3
AP003361
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 4/9.
ACCESSION
AP003361 BA000017
VERSION
AP003361.2 GI:14246761
KEYWORDS
SOURCE
Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Matsunaru,H., Maruyama,A., Murakami,H., Hoshoyama,A.,
Mizutani-U,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
TITLE
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL
Lancet 357 (9264), 1225-1240 (2001)
MEDLINE
21311952
REFERENCE
2 (bases 1 to 347235)
AUTHORS
Ohta,T.
TITLE
Direct Submission
JOURNAL
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-noodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT
On May 29, 2001 this sequence version replaced gi:13875305.
FEATURES
source
location/Qualifiers
1..347235
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
35..1021
/gene="SAV0992"
35..1021
/gene="SAV0992"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter ATP-binding protein
homologue"
/protein_id="BAB57154.1"
/db_xref="GI:14246762"
/translation="MNNVLEVKDLETSLKINNEWLATVENISPELSKGEVLGIVGES
GCKSTLSKSIKLLPKSKISLSSGEVIFDGRIDTLENEKOLLIDRGNDIAMIFQPM
TALNPFYIKQVLVESIKSHKISKKEANNLAKDLLKKVIAQDELNSYPHQLSGG
MRGVANALNSCKLLIADPTLDVITQAILDKELQREYOMAMIMTHDLS
VVAEFCDKVIMYAGQIVFEGGIEILHNPKHYTKLLSTIPKLKSGQRLETTLEGI
VPSIQAFHNKCRFRANRCKNKLDCNQNQSPKMHVCEDEVIVRCHLYKNEYKEI"
1024..2004
/gene="appf"
1024..2004
/gene="appf"
/notes="SAV0993"
/codon_start=1

gene
CDS

gene
CDS

gene
CDS

gene
CDS

```

```

/transl_table=11
/product="oligopeptide transport system ATP-binding
protein AppF homologue"
/protein_id="BAB57155.1"
/db_xref="GI:14246763"
/translation="MENILEVNOIKKYIKTKLOKTOYVKAVDVSVFSIKKGOTFG
LYSGGCGKSTLAKVIIRLEDATSGSIIVNGEDITRLOGKKLSPQOYQMIPODPYA
SLNPNQWGDIISEPIILNKKLPKEIKKEVLYLLKCVGLSEDAFYKHAHESGGQKQ
RVGTARALARPSSLIVADPEVSALDVSQSVLMLKDQEFKLSYLFIAHDLUSVK
HISDVIYGMVLGHIVIAEDSKBIYENPKPYTKALISSIPQIDKHNNRIILKGLPFS
PSNPQCGPHTRCPIAKDKCKENIPQLKIDICNEHQVACFYVDKVGDLNA"
1997..2959
/gene="oppB"
1997..2959
/gene="oppB"
1997..2959
/gene="oppB"
/notes="SAV0994"
/codon_start=1
/transl_table=11
/product="probable oligopeptide transport system permease
protein"
/protein_id="BAB57156.1"
/db_xref="GI:14246764"
/translation="MLKLKRLGLMPLLLISIVFSLAIQPGDPSPDLQNGKIK
QEAINAQREKLGINDSIPTHQYIRWVNHVIRHGDLAGESIKYKRPVIDVIEIRIPNTILG
AMSLITVIISFALGITSGRYSVSLDTVQVFNLMALPFIAGVFAIFSPFELQ
WPFQGSVDINLKEGTFEYMSKIYHTFLPALTLGLLSTAGYIQYLRNDIENSCKDY
VLTASKGLSNKKIYNKHILRNSLIPITFLFGADIVSLGAVITETIFSYNGIKLFL
LSEVIGQDIPMLMALTLFFSLGLGNLSIDITFGIDPRIRSN"
2971..3852
/gene="SAV0995"
2971..3852
/gene="SAV0995"
/notes="similar to oligopeptide transport system permease
protein oppC"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57157.1"
/db_xref="GI:14246765"
/translation="MQNKSPPKIALSRFTHNKIAMSIFILLIITIYSIIAPLIAP
FVNQDQDLNTKGMTAQNILGDSGRDNFSLIYAGRIISLSITGITSTIGMLLIGIT
FVIVSGYGGIVDTLMTITFVFLPFIPLFAIVLNAALGKIKNPGYSAIILVLVII
VLSWGGIARLVGRKVLQERNEYFAAKSIGTPTPKILKHELLPILSVIIVQATLFL
AGMIVVESGLSFLQGISKAIPSGNMLSDAQEGSVTSGRKPWTWMPPALMITLTLSI
NFVGSGLDAFNPGR"
complement(3894..4883)
/gene="trpS"
complement(3894..4883)
/gene="trpS"
/notes="SAV0996"
/codon_start=1
/transl_table=11
/product="tryptophanyl-tRNA synthetase"
/protein_id="BAB57158.1"
/db_xref="GI:14246766"
/translation="METLFSGIQPSGIPTIGNYIGALKQFVDVNDYDCYFCIVDQHA
ITMPDRLKRLKOTROLAAIYLASGIDPDKATLFIQSEVPAHQAGWMLTIASVGEI
ERMTQYKQAKAVGEPAGLLIYPLMAADIVLYNTNIVPVGDOKQKIELTRNLVD
RNSRYNDVYKPIRMPKVGVRVMSLQDPTKMSKSDNKNFISLIDEPNVAKKI
KSAVDSGDIILKFRDNKPGITNLISYAGLTDMPDKIEAKYEGEGYKFGKGLDAEI
VKAFIVFQERKEYSFNSDKLDDILDQGRDKAHKVSFTVKRMEKAMGLGRK"
5177..5572
/gene="SAV0997"
5177..5572
/gene="SAV0997"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57159.1"
/db_xref="GI:14246767"
/translation="MVTILFTSPSCTSCRKAKWLQEHDPYTERNIFSEHLLIDRIKQ
ILKMTEDGTETIIRTSKTYQKLNVDISLPLQDIYSITQDNPGLLRPIILDNKRLQ
VGYNEDEIRFLPKRVTFQLQEAQRWD"

```

```

gene      5943..5662
/ gene="SAV0998"
CDS       5943..5662
/ gene="SAV0998"
/ note="similar to negative regulator of genetic competence
MeCa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57160.1"
/db_xref="GI:1424678"
/translation="MRIERYDDTTFVKLFITYSDIEARGFREDLWNRKRGBEFTWSM
MEIINEEDFVEEGLPIQVHAFEGEVITISKNEEDMMNSDDATOFQEQVEL
LAQTLEGDELEELFEQRTKEAOKSRKSNKTRIIIVKFNLDLEDVINYAHS
NPITFEEDLLVMVDGYVYAVHFDSDVDQEVINDSYSQLERAYPTDRTVEYLNDYA
KIMSHNTAQVRRYPETTE"
6783..7769
/ gene="SAV0999"
CDS       6783..7769
/ gene="SAV0999"
/ note="similar to transcription factor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57161.1"
/db_xref="GI:1424679"
/translation="MLVALNEEKERVLTATLRKTOYFPCVCGKQVILKRLKVISHF
AKHLAEOKCFNNEETIKHYKSLILIAOMIOGCKVELEPELKEIKOIPDILINKYV
IELQSPPIYQIQLQRTGELKKMGYKVSWLLNDVYCHNKYKNHPSFLFIPNIRKL
HTFNLEKKQIMFOIOIQLGGHKYVAEKNAKIIELFNEAPCDYHAYILSKFPAINQY
IKYKQNSVLEPTLSAMYQLQTDQEVYNYGYIFPEQIYIENHPTEWQLQVLDLX
NKSILVNDNLNFKLKKFIVALESKTAIEIKLINNYLNICSDRGNDVQILF"
7817..9625
/ gene="SAV1000"
CDS       7817..9625
/ gene="SAV1000"
/codon_start=1
/transl_table=11
/product="thimet oligopeptidase homologue"
/protein_id="BAB57162.1"
/db_xref="GI:14246770"
/translation="MSQLSREGEKRYPEYVWDLTIFTKDDEAFEAKEVENELGK
EQFGHGHSDEATLYNALBEDLTGKLEKVVYVAHLKQDDPTNDKYPGMESSRAHQ
LIKFKSGAVFPEIQLDIEDXISFVNSYDKLQKFAFDLKLINERPHILDAETEK
LUTEADALSPNSVYGMFNADIVFEDAIKDGNAHPLTQGTPIKYLESDDRKLRES
APRNYKAVGAHNTLGATLAGEVKNVENARTHNKTAAREKALSNHNPENVYDNLV
KTVKYLPIHRYELRKELGLDGLKVDLYTPIKDIKEMPYREPAKFWMLKALEP
MGEYLNVVKEGLNNRVYDVENKGRSGYSSGAHLNPPILLNWSNTISDLTYLVE
EFGHSYFSKFKQPSNSDYTFVAEVASTCNEALLSDYMDKHLDDERKLLLNQE
LERFRATFQTMFAEFHKKIHAEAGEPLTPTRMNEYAKLNKLYFGDSVETEDI
SKWSRIPHFMNYVYQYATGYSAQSLSHQLTGEKPAVDRYINEFLKKGSSNYPY
EILKNAGVDMTTPEIEQACEVPEOKLNAPEKLMA"
complement(10085..10639)
/ gene="yjbM"
CDS       complement(10085..10639)
/ gene="yjbM"
/ note="SAV1001"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="BAB57163.1"
/db_xref="GI:14246771"
/translation="MKOLSKVWSKTEKIYTYCYNEIIPKRDITBSMTCDCIONAG
IDLEVFKDLOKSKLTESLKDILHAREMEIEQAPSLVFFSEDVHEGLKVEGLYPYH
IYYIYNELMGPIEKNLPKLEFYIQOQLYTMSELLTYEWPEKLLINKELKKAIAQ
QKLEKLPDGDFFWKSMPKIKSK"
complement(10575..10901)
/ gene="SAV1002"

```

Alignment Scores:

```

Pred. No.:      3,57e-127      Length:      347235
Score:          2274.00      Matches:      455
Percent Similarity: 100.00%      Conservative: 0

```

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match:          100.00%      Indels: 0
DB:                   1           Gaps: 0
US-09-943-108a-2 (1-455) x AP003361 (1-347235)

QY 1 MetalapheluglyleuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 256730 ATGGCAITTTGAAGGCTTATCAGACGCTGCAAGCGACGATGCAAAAAATCGTGTAAG 256789
QY 21 GlyLysLeuThrGluAlaAspLysLeuMetArgGluValArgLeuAlaLeuPhe 40
Db 256790 GGTAAACTTACTGAAGCTGATATAAGATAATATGCGTGAAGTAAGATTACGCTTATT 256849
QY 41 GluAlaAspValAsnPheLysValLysGluPheLeuLysThrValSerGluArgAla 60
Db 256850 GAGGCTGACGTAAACATTTAAAGTGGTAAAGAAATTTATTTAAACAGTATCAGACGCGCA 256909
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysLeuValGln 80
Db 256910 TTAGGTTCCGATGTAATGCAATCATTAAACACGCGCAACAAGTTATTAAATAGTTCAA 256969
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256970 GATGAATTAAGAGTTGATGGGTGGAGAAAATACATCGATTATATATGTCAAAATAAACCA 257029
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 257030 CTTACTGTTGTATGATGTTGGTTTACAAGTCTGTTAAACACACAACTCGACGTAAA 257089
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 257090 TTAGCATTTATTGATGCGTAAAAAATACAAACAAAAAACCCTATGTTAGTTCAGCAGATATT 257149
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 257150 TATCTCCAGCAGCGATAAATCAATTACACAGTAGGAGAAACAAATTAATATCTCTGTA 257209
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 257210 TACAGTGAAGGAGATCAAGTAAAGCCACAAACAAATTTAACTAATGCATTAAACATGCT 257269
QY 181 LysGluGluHisLeuAspPheValIleleAspThrAlaGlyArgLeuHisleAspGlu 200
Db 257270 AAAGAAGAACATTTAGACTTTTGTATTCATTATGATACAGCAGCTCGATTACACATCGATGA 257329
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLeuLeaLysProAsnGluIleMetLeu 220
Db 257330 GCATTGATGAAGAAATTAAGAAGAAAGTAAAGAAATTCGTAACCAACCAAGAAATTAATGTTA 257389
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 257390 GITGTCGATTTCAATGACGGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTTGACGATCAA 257449
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 257450 CTTGATGTCACAGGTGTTACCTTAACATAAATTAGATGTTGATACACGTGGTGGTGCACCT 257509
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 257510 TTATCTATTCTGTCGTGACACAAAAACCAATTAATTAATTTGTTGGTATGAGTGAAAGTTA 257569
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 257570 GATGCTTTAGAGCTATTCCATCTCTGAACTGATGATGATGATGATGATGATGATGATGAT 257629
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 257630 GTGTTAAAGTTTAAATGAAAAAGCGCAACAAAGATGTGATGCAAGAAAAAACAAGATTTA 257689
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
Db 257690 GAGAAAAAGATGCGTGATGATCATCGTTACTTTTACTTTTACTTTTACTTTTACTTTTACTTTT 257749

```



```

Db 27184 TTATCTATTCTGTTCCGTTGACACAAAACCAATTAAATTTTGTTGTCAGTGAAGAGTTA 27243
QY 281 ASpGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||
Db 27244 GATGGTTTAGAGCTATTCCATCCCTGAACSTATGGCATCACGTATTTAGGTATGGGTGAT 27303
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||
Db 27304 GTGTTAAGTTTAAATGAAAAAGCCACACAGATGTGGATCAAGAAAAAGCAAGATTTA 27363
QY 321 GluLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||
Db 27364 GAGAAAAAGATCGTGAATCATCGTTTACCTTTAGATGATTTTATAGAACAACTTGATCAG 27423
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||
Db 27424 GTGAAAAATCTAGGACCACTGGATGATATTATGAAAAATGATCCAGGTATGAATAAGT 27483
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||
Db 27484 AARGGCTAGATAAGCTTAATATGAGTGAAGCAAAATTCATCATATTAAGCGATTATC 27543
QY 381 GluSerMetThrProAlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLys 400
|||||
Db 27544 CAGTCATGACGCGCTGAAAGAAACAATCCAGACACATTCGAATGTATCAGCTAAAAAG 27603
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||
Db 27604 CGTATTGCTTAAAGGCTGGTGGTTCATTACAAGAGTCAATCGTTGATGAACAAATTT 27663
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLys 440
|||||
Db 27664 AACGATATGAGAAATGATGAACATTCACCTGGTGGCGGTAAAGGTAAAGGTAA 27723
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||
Db 27724 CGCAATCAAAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 27768

RESULT 5
AX416333/c
LOCUS AX416333 2256 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 3324 from Patent WO0228891.
ACCESSION AX416333
VERSION AX416333.1 GI:21448790
KEYWORDS
SOURCE Listeria monocytogenes ATCC 19115.
ORGANISM Listeria monocytogenes ATCC 19115
REFERENCE
1
AUTHORS Glaser,P. and Kunst,F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 3324 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
source
location/Qualifiers
1..2256
/organism="Listeria monocytogenes ATCC 19115"
/ob_xref="taxon:176281"
BASE COUNT 533 a 471 c 420 g 832 t
ORIGIN

Alignment Scores:
Pred. No.: 5,36e-96 Length: 2256
Score: 1719.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best Local Similarity: 72.75% Mismatches: 55
Query Match: 75.62% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX416333 (1-2256)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgLys 20
|||||
Db 1666 ATGGCATTTGAGGACTAGCTGGAAGACTCCAGAAACAACTGATGATTTGAAGCAATATC 1607

```

```

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||
Db 1606 GGAAGAGTAAACGAGCTGACGTAAAGAAATGATCGCTGAAGTCTGCTGTCTACTT 1547
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
|||||
Db 1546 GAAGCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACAGTAAGCGAAGTCT 1487
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||
Db 1486 GTGCGCGCAGCGTTATGAAAGCCTAAACCGCGCCAAAGTTATCAAAATCGTTCAA 1427
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||
Db 1426 GAAGACTAACAGCTTAATGGCGGAGAGAAACRAAATCGGAACAGCGACGCCCA 1367
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||
Db 1366 CCAACCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1307
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
|||||
Db 1306 CTTGCTAAATTTATACGCAAAATATATAACCGTAAACCACTTACTTGCAGCAGATAT 1247
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||
Db 1246 TACCGACCTGCGCGCAATCAAAATATAGAACACACTTGCAGCAATATGATATGCCGTA 1187
QY 161 TyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeuLysHisAla 180
|||||
Db 1186 TTTTCTCTAGGAGACCAAGTAGCCGGTAGAAATCGCAACCAAGCTATTGCTAAAGCA 1127
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||
Db 1126 AAAGAGAAACATTTAGATTATGATATGATATGATATGATATGATATGATATGATAT 1067
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
|||||
Db 1066 ACCGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
|||||
Db 1006 GTTGTGATTCATGACGTGGCAAGATGCTGTCACAGTGGCGCAAGCTTCAACGAACAA 947
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
|||||
Db 946 TTAGAAATTACAGGTGTCGCTCAACCAATTAGACGGGTGATACACGTGGTGGTCAGCG 887
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlyLysLeu 280
|||||
Db 886 CTTTCGATTCGTTCCGTCACCTGGCAACCAATTAATTTATCGCAGCTGGTGAAGAAATG 827
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||
Db 826 GAAGCTCTAGAAACTTCCATCCAGATGATGGCTTCAGAAATTCGCGCATGGCGCATG 767
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||
Db 766 GTGCTTTCCTGATTGAAAAAGCGCAACCGATGTAGATACAGAAAAAATGAAGCTATG 707
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||
Db 706 GAACAAAAATGAAGACACACAGCATGACGCTAGATGACTTCTTAGAACAGTTACAAACA 647
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||
Db 646 GTAAAAACAATGGGACCACTAGATGAAGTACTTAAAAATGATGCCGCGGCAACAAATG 587
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||
Db 586 AAAGCGCTGACAACTGAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 527

```


/note="similar to hypothetical proteins"

```

/codon_start=1
/transl_table=11
/protein_id="CAC99773.1"
/db_xref="GI:16411147"
/db_xref="SP:SPREMBL:Q8Y6J1"
/translators="MKITSISVQOKKERYNIFDEKYNFSVDEEVLYARYLQMKGVYL
TEADIEIKQADWVRKGLNKAIFLSHRVSEKEDYLRKQEMEPFALDEILKLAN
MDYINDEFELVYKCIKTLTKGRTYELRELVKELREIISRVIEYSDEAOLENA
KQOAMTKMRNKSARKMLQOKIITDILQKGFSELAKAATAEATSELDVADEAELIQ
KOIEKTKRKNRYKPSIAKOKIITSLMOKGFSYDTIQSVLTENEISFEEB"
complement(3872..3877)
/gene="lmo1692"
RBS
complement(3872..3877)
/gene="lmo1692"
RBS
complement(4680..4685)
/gene="lmo1693"
gene
join(4757..5673,5681..5708)
/gene="lmo1694"
RBS
4757..4762
/gene="lmo1694"
CDS
4771..5673
/gene="lmo1694"
/note="similar to CDP-abequose synthase"
/codon_start=1
/transl_table=11
/protein_id="CAC99772.1"
/db_xref="GI:16411148"
/db_xref="SP:SPREMBL:Q8Y6J0"
/translators="MHLLTGATGFIGDHLVHELEKSDHELYLTRQKLKRNANVHI
EWLNEDPLENDLPVDVGINLAGAGLMDKWTYDRKKVIVNSRIEATSAIISVKKM
KSKPLWNSALGAYTSKSTIYLDTEENTYADNFTGKTYVENEKTSASADIGIV
VYAREGLVNGSGSPFPEKLFQYITGREGNGRQWYSHIHVDDVVAALIFDHDQ
LVNGVTAHPVOEKKFAERKMKHPKTPVPKKILKILGERAMTILDSRAPY
EKLSNHFEEFFTEQLDLDLD"
5681..5708
/gene="lmo1694"
terminator
complement(join(5682..8291,8299..8304))
/gene="lmo1695"
gene
complement(5682..5709)
/gene="lmo1695"
terminator
complement(5694..8291)
CDS
/gene="lmo1695"
/note="similar to putative membrane proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99773.1"
/db_xref="GI:16411149"
/db_xref="SP:SPREMBL:Q8Y6I9"
/translators="MKEKLMQYAWFQKNVVKIVITFVMAFVIFELINATGIDY
PSIKENLTSQSPQIETIMPTVGLIATVPMIYDVIYKLLPGKFSFVSIASGRTWT
FTNGGGVGLASLRASPYGNASHKEILLAIKIALFLVSLGSLTYCLVSLAILIP
GCFADHFVNYFWLLAGLGYFPILFTITRWKSKSLFVLDPIKRELITLITASLLEWGFAP
GCFALITGMEGPDIFKVPFLFIASVIGIASMPGGVGTEDVMVILGSLQVSOE
LALAMFLYRIFYIIPVYVGLLFFVQAGKNVDLEGLFLFLQKVAHRLFIWY
GSGILLILSSAVNAIYHVPFLYKIMPNFLFTSQITIVAFGLILGLARFTECKTK
AVIITVILGCAFNILARVSMKQAIPLGVLLCLLARNERYREKIVYTWKSVIID
SLITFVCLAGVIVGLYNSPNLKHSEIPLDXLRASEHLWLVGVGVYVAYVSLVIY
ITLITKEKGSPEAVKVFREHLARKNGENVSHFTLRKLLFWAREGEVLFSTRIL
DKWIVNGPTNGMKKEALEEVMNADRFGYRPVFEYEVGMTIMPYLHDHGFDFIKG
DEGIVNONTMSSKKKGERALMKNLREGYTFEIIPEFPNHDITWTLRAVSDWLD
PREEKGSFLGFTDYITLQAPATAIKNGEITVIGVFSNMPSTDEMTSIDLKRYSKA
PSGIMDPLFINLPEKEDQFQFNAGMAPLANVGSKYAFILGERLAGIVRYSGVY
GFKGLRNEKSKYVTEWEQKFAFRKRSSATFMTQLMLLVGKKRPLANSQVVDLPELE
BETKRPDSE"
complement(8299..8304)
RBS
/gene="lmo1695"
gene
complement(8318..9007)
/gene="lmo1696"
CDS
complement(8318..8992)
/gene="lmo1696"
/note="similar to unknown proteins"

```

```

Alignment Scores:
Pred. No.: 2,22e-93 Length: 347050
Score: 1717.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best Local Similarity: 72.75% Mismatches: 55
Query Match: 75.53% Indels: 5
DB: 1 Gaps: 1

US-09-943-108a-2 (1-455) x AL591981 (1-347050)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 122116 ATGGCATTGAAGGACTAGCTGGAGACTCCAAAGAAACATGACAAATAATCGCGGTAA 122057
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 122056 GGGAGATAAAGCAAGCTGACGTAAAGAAATGATCGGTGAAGTTCGCTTGTCTACTA 121997
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121996 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACAGTARGCAACGTGT 121937
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 121936 GTCGCGCGGACGTTATGAAGAAGCTTAACACCGCGCAACAGATTATCAAAATCGTTCAA 121877
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 121876 GAAGCAATACAAAGCTTAATGGCGGAGAGAAGCAAAATCGGAACAGCGGACCGCCG 121817
QY 101 ProThrValValMetMetValGlyLeuGlnGlyValaGlyLysThrThrAlaGlyLys 120
Db 121816 CCACCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121757
QY 121 LeuAlaLeuLeuMetArgLysLysTrpAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 121756 CTTGCTAATTTATACGCAAAAATATAACCGTAACCACTTACTTGTGCGACGAGATATT 121697
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 121696 TATGACCTGCCGCAATCAAAACAATTAAGAACCCTTGGCAAGCAATTAAGATATGCGGTA 121637
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 121636 TTTTCCCTAGGACACCAAGTAGCCCGTCGAAATCCCAACAAACGCTATTGCTTAAGCA 121577
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 121576 AAGAAGAACATTTAGATTATGATTATGATTATGATTATGATTATGATTATGATTATGATT 121517
QY 201 AlaLeuMetAspGluLeuLysGluValLysGluIleAlaLysProAsnGluLeuMetLeu 220
Db 121516 ACGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121457
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaIleAspPheAspAspGln 240
Db 121456 GTAGTCGATTCCATGACCGGCGCAAGTGTCTCAAGTGGCGCAAGGCTTCAACGAACAA 121397
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 121396 TTAGAATTACAGGTGTCGTATTACGAACTAGATGGTGAACGTCACGTCGTGGTGGCGG 121337
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 121336 CTTTCGATTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 121277
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 121276 GAACCCCTTGAAACCTTCATCCAGATCGTATGCTTCCAGATTCCTCGCATGGGGAT 121217
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspIleGlnLysAlaLysAspLeu 320
Db 121216 GTGCTTTCCCTGATTGAAAAAGCGCAACCGGATGTAGATACAGAAAAAATGAAAGCTATG 121157

```

```

QY 321 GluLysMetArgGluSerPheThrLeuAspPheLeuGluGlnLeuAspGln 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121156 GACAAATAAGAGACACAGCATGACGTAGATGACTTCTAGACAATTACACAA 121097

QY 341 ValLysAsnLeuGlyProLeuAspAspPlexMetLysMetIleProGlyMetAsnLysMet 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121096 GTAAACAAATGGGACCACTAGACAACTACTTAAATGATGCCAGGGCAACAAATG 121037

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121036 AAAGCCCTGCACATGACATGACATGATTAACAACTGATGCTATGACCAATTATC 120977

QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120976 AAATCAATGACCAAAATGAAGACACACACACCATCATCATCATGACAGTAGAGAAA 120917

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120916 CGGATTGCACGTGATGACGACGTCGCGTTCAAGAAATTAACCGTCTCTCTAAACAATTT 120857

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120856 GCTGAATGAAATAATGATGAGCAATGATGCTGTCGCGCAAGCAAAAAGGCAAA 120797

QY 441 ArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120796 -----AACCATTCGCGCAATTTCAAGATGCCAATTT 120767

RESULT 7
AL596170/c 313450 bp DNA linear BCT 06-JUN-2002
LOCUS
DEFINITION Listeria innocua Clp11262 complete genome, segment 8/12.
ACCESSION AL596170 AL592022
VERSION AL596170.1 GI:16414292
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussurget, O., Entian, K.D., Fsihi, H., Portillo, P.G., Garrido, P.,
Gautier, L., Goebel, M., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F.,
Kurapat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E.,
Nedjari, H., Nordisk, G., Novella, S., de Pablos, B., Perez-Diaz, J.C.,
Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N.,
Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
TITLE
MEDLINE 21537279
PUBMED 11679669
REFERENCE
2 (bases 1 to 313450)
AUTHORS Glaser, P., Frangeul, L. and Rusniok, C.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
COMMENT E-mail: pglaser@pasteur.fr
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
FEATURES
Location/Qualifiers
Source
1..313450
/organism="Listeria innocua"
/strain="Clp11262"
/db_xref="taxon:1642"
complement(287..1444)
/gene="lin1790"
complement(287..1444)
/gene="lin1790"
/note="similar to transmembrane transport proteins"

```

```

/codon_start=1
/transl_table=11
/protein_id="CAC97021.1"
/db_xref="GI:16414293"
/db_xref="SPTREML:Q92AX8"
/translation="MGAKGFMTLTVVAISGLSQGVLLPIAIIIEKGVSGVINGF
IGATIGLVGVSPFIEAPLHKYKPIILVGGGVAIFAIPIWENLYFWFLRL
HIGDHMLHFSOTWIGANDPSKGRNMALYGLFFSVGPAIGPOLVNAKYANLPF
FISGLIYLIANSLVFIRNDFVEGEKAVIRKISFNGSKRSVGFKAWMAMIPPLYG
ILETGLNATFFPVGLRDGLDTMMIASSFSVGTIIFQVPIGLISDKTRDKMLPLL
TGAGAVVFLFAFKIDIFFVIFLIGLLGSYLSGLSYMTDLTFLLEPAGNLY
GFFSLGSIIGPSATGMIGTIFGNQIYFFWAGILLGLCLLALGGKKEAH"
trna
/product="tRNA-Glu"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 69.21"
complement(1585..1675)
trna
/product="tRNA-Ser"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 63.76"
complement(1712..1784)
trna
/product="tRNA-Asn"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 85.46"
complement(1791..1864)
trna
/product="tRNA-Ile"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 92.91"
complement(1891..1961)
trna
/product="tRNA-Gly"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 73.1"
complement(1984..2056)
trna
/product="tRNA-His"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 72.11"
complement(2083..2155)
trna
/product="tRNA-Phe"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 79.39"
complement(2160..2235)
trna
/product="tRNA-Asp"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 73.32"
complement(2240..2313)
trna
/product="tRNA-Met"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 78.6"
complement(2356..2445)
trna
/product="tRNA-Ser"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 66.89"
complement(2468..2541)
trna
/product="tRNA-Met"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 90.88"
complement(2553..2626)
trna
/product="tRNA-Met"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 77.07"
complement(2681..2753)
trna
/product="tRNA-Ala"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 84.64"
complement(2773..2846)
trna
/product="tRNA-Pro"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 86.85"
complement(2857..2930)
trna
/product="tRNA-Arg"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 74.21"
complement(2940..3025)
trna
/product="tRNA-Leu"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 66.88"
complement(3040..3114)
trna
/product="tRNA-Gly"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 87.88"
complement(3130..3211)
trna
/product="tRNA-Leu"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 69.82"
complement(3217..3289)
trna
/product="tRNA-Lys"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 88.37"
complement(3330..3405)
trna
/product="tRNA-Thr"
/note="tRNA-Scan-SE vs 1.3 result - Cove score = 95.38"
complement(3414..3486)

```



```

Db 138781 TTAGAATAACCGCGCTGTATTAACAAAAATTAGACGGTGATACACGTCGTGGGCGACGA 138722
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138721 CTTTCCATCGTTCACTACAGAAACCAATCAATTCGTGCTACCGGTGAATAATG 138662
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||| :||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138661 GAAGCAATCGAAACGTTCCATCCCATGCTGATGCTTCAAGAAATTCCTGGCATGGTAT 138602
QY 301 ValLeuSerLeuIleGluLysAlaGlnbNAspValAspGlnGluLysAlaLysAspLeu 320
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138601 GTACTTCTCTTATTGAAAAGACACAACTAGTAGATGACGAAAAATGAAGCTATG 138542
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138541 GAACAAAAATGAAGACACACAGATGACCTTAGATGACTTCTGGACCAATTCACAA 138482
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138481 GTAAAACAAAATGGGACCACTAGTAGACTACTAAAAATGATGCCAGGGGCAACAAATG 138422
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
||||| ||||||| ||| :||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138421 AAAGCCTCGACACATGCAAGTAGATGATAACAACTCGGTACACATCGAAGCGATATT 138362
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138361 AAATCCATGACCAAAACGAAAGAAATATATCCGGACATCATCAATCGGACGAGAGAAA 138302
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138301 CGAATTCGTGCGTGAAGCGGACGCCCAATTCAAGAAATCAATCGCCTCTAAACAATTT 138242
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 138241 GCTGAATGAAAAAATGATGAAGCAAAATGACTGGTGGAGGAAAGGTAAAGAAAGTAAA 138182
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
||||| ||| :|||||
Db 138181 -----AATCCATTCGGCAATTTCAAAATGCCATTT 138152
RESULT 8
AX413016
LOCUS AX413016 319630 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 7 from Patent WO0228891.
ACCESSION AX413016
VERSION AX413016.1 GI:21445474
KEYWORDS
SOURCE
  Listeria innocua.
  Listeria innocua
  Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
  1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 7 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
  Location/Qualifiers
    1..319630
     /organism="Listeria innocua"
     /db_xref="taxon:1642"
BASE COUNT 105207 a 55428 c 66726 g 92263 t 6 others
ORIGIN

```

Alignment Scores:

```

Pred. No.: 7,12e-93 Length: 319630
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

```

US-09-943-108a-2 (1-455) x AX413016 (1-319630)

```

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 187924 ATGGCAATTGAGGACTACCTGGAGAGACTCCAGAAACAATGACAAAATTCGCGTAAA 187983
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLysPhe 40
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 187984 GGAAGAAATAAAGCAAGCTGACGTAAGAAATGATGGTGAAGTTCGTCCTTTACTA 188043
QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188044 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACACAGTAAGCGAAGCGCA 188103
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlnGlnValIleLysIleValGln 80
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188104 GTTGGCGGAGAGTATGAAAGAGCTACACCCGCTCAACAAGTATTAATAATGTTCAA 188163
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188164 GAAGAATTAAACAGCCTCATGGCGGCGGAGAAAGTAAATCGGAACAGCTGATGCGCGC 188223
QY 101 ProThrValValMetMetValGlyLeuGlnGlyValGlyLysThrThrAlaGlyLys 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188224 CCAACCGCTTATGATGTTGAGGAGCTGTTCAAGGAGCTGGTAAACACACCCTTCAGGAAA 188283
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188284 CTCGCTAATTTATGCTAAAAAATATATCGTAAACCTTTACTAGTCGACGAGATATT 188343
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188344 TACCGACCTGCGCAATCAACAATTAAGAACACTTGGCAACAATAGATATGCGAGTA 188403
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188404 TTCTCTTAGCGCATCAAGTAAGCCAGTAGAATAATCGCAACCAAGCTATCGCTAAAGCA 188463
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188464 AAAGAAGAACAATTAGATTATGTCATTATCGATACAGCTGGTGTCTTCATATGACGAA 188523
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188524 ACCTCGATGCGCAATTAACAAAGTGAAGAAATTCGCTACGCCAACTGAAATTTTACTT 188583
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188584 GTAGTTGATTCAATGACTGGGCAAGACCGCATAAATGTGCCCCAAAGCTTCACGACAA 188643
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188644 TTAGAANTTACCGCGCTGTGTTATTAACAAATTAGACGGTGATACACGTTGGTGGGCA 188703
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188704 CTTTCCATCCGTTCACTGACAGAAACCAATCAAAATTCGTTCTACCGGTGAAAAATG 188763
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188764 GAAGCAATCGAAGCTTCCATCCCATCGATCGTATGCTTCAAGAAATTCGGSCTATGGTGAT 188823
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188824 GTACTTCTCTTATTGAAAAGACACAACTGATGTAGATGCAGAAAAATGAAGCTATG 188883
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188884 GAACAAAAAATGAAGACACACATGACCTTAGATGACTTCTTGGCAATTTGCAACA 188943
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 188944 GTAAACAAATGGGACCACTAGATGAATCTACTAAAAATGATCCAGGGGCAACAAATG 189003

```

```

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLeuAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189004 AAAGCCTCGACAACTGACAGTAGATGATRAACAACACTCGGTACATCGAAGCAATTT 189063

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189064 AAATCCATGACCAAAAACGAAAAAGATAATCCGGACATCATCAATCGCAGAGAAGAAA 189123

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189124 CGAATTCGTGGAAAGCGAGCCCAATTCAAAGAAATCAATCGCCTCTTAACAAATTT 189183

QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189184 GCTGAATGAAAAAATGATGAGCAATGACTGGTGGAGGAAAAGCTPAAGAAAGGTAAA 189243

QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 189244 -----AATCCATTGCGCAATTTCAAAAATGCCATTT 189273

RESULT 9
LOCUS AX417046 349980 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 4037 from Patent WO0228891.
ACCESSION AX417046
VERSION AX417046.1 GI:21449656
KEYWORDS
SOURCE
ORGANISM
Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1
AUTHORS
Glaser, P. and Kunst, F.
TITLE
Listeria innocua genome and applications
JOURNAL
Patent: WO 0228891-A 4037 11-APR-2002;
PASTEUR Institut (FR)
FEATURES
Location/Qualifiers
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/notes="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"
BASE COUNT 101055 a 72969 c 60688 g 115268 t
ORIGIN

Alignment Scores:
Pred. No.: 7,89e-93 Length: 349980
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 6 Gaps: 1

US-09-943-108a-2 (1-455) x AX417046 (1-349980)

QY 1 MetaLapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgLys 20
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129501 AFGGCATTTGAAGACTAGCTGGAAGACTCCCAAGAAACATGAACAAATTCGGGTAAA 129442

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129441 GGGAAAGTAAACGAAGCTGACGTAAAGAAATGATGCGTGAAGTTCGTGCTTTACTA 129382

QY 41 GlnAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129381 GAAGCCGATGTTAACTTAAAGTCGTTAAACATTTATTAAAAACAGTAAGCAAGCGTGA 129322

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129321 GTTGGCGGAGCTAATGAAGAAGCTTAACACCGGTACACAGTAATTTAAATCGTTCAA 129262

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129261 GAAGAATTAACAAGCTCTATGGGGGGAAGAAATGAAATCGGAACAGCTGATCGCGCC 129202

QY 101 ProThrValValMetMetValGlyLeuGlnGlyValAsnLysIleAspIleLys 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129201 CCAACCGCTTATATGATGTAGGTTCACAGGAGCTGGTAAACCAACCACTTCAGGAAA 129142

QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129141 CTCGTAAATTTATACGTAAATAATAATACGTAAACCTTTACTAGTCGACAGCATATT 129082

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129081 TACCGACTCGACCAATCAAAACAATTAAGAACACTTGGCAAAACAATTAAGATATGCCAGTA 129022

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 129021 TTTTCTCTAGCGCATCAAGTAAGCCCACTAGTAATCGCAACAAGCTATCGCTAAAGCA 128962

QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128961 AAAGAAGACATTTAGATTATGCTATTCATGATACAGCTGGTCTTTCATATCGACGAA 128902

QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleAlaLysProAsnGlnIleMetLeu 220
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128901 ACTCTGATGAGCAATTAACAACAATTAAGAAATCGCTACGCCCACTGAAATTTTACTT 128842

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128841 GTAGTTGATTCATGACTGGCGCAAGACGACGTAATGTGGCCCAAGCTTCAACGAACAA 128782

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128781 TTAGAAATTAACCGCGCTGTGTTATTAACAAAATTAAGCGGTGATACACAGCTGGTGGCGCAGCA 128722

QY 261 LeuSerIleArgSerValThrCluLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128721 CTTTCCATCGTTCAGTCACAGAGAAAACCAATCAAAATTCGTTGCTACCGGTGAAAAAATG 128662

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128661 GAACCAATCGAAGCTTCCATCCGATCGTATGCTTCAGAAATTCCTCGSCATGGGTGAT 128602

QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysLeu 320
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128601 GTACTTCTCTCTATTGAAAAAGCACAAACTGATGATGACAGAAAAAATGAAAGCATG 128542

QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128541 GAACAAAAATGAAGAACAGACAGCATGACCTTAGATGACTTCTTGACCACTGCAACAAA 128482

QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128481 GTAAACAAATGGGACCACCTAGATGAACACTACTAAATAATGATGCCAGGGGCAACAAATG 128422

QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128421 AAAGCCTCGACAACTGCAAGTAGATGATAAACAACATCGGTACATCGAAGCATATT 128362

QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128361 AAATCCATGACCAAAAACGAAAAAATAATCCGGACATCATCAATCGCAGCAAGAAAA 128302

QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 128301 CGAATTCGTGGAAAGCGAGCCCAATTCAGAAATTAATCAATCGCTCTTAACAAATTT 128242

QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```


Qy	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
Db	1491	CTTCTATTTCGGCGCTGCACACACCCCAATAGTTTGAGGTTTGGCGGAAGACTT	1550
Qy	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	1551	GATCGCTAGAGCCGTTCCATCCTTGAAGCATGGCATCAAGGATTCTCGCGATGGCGAC	1610
Qy	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu	320
Db	1611	GTCTGACATTGATTGAAAAAGCAGCGCGTGTGATGAAGACAAAGCCAAAGAGCTG	1670
Qy	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
Db	1671	GAACAAAAATCAGAACCATGAGCTTCACATTGGACGATTTCTGGACGACCTCGGGCAA	1730
Qy	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	1731	GTGAGAAACATGGGCGCCTTGATGAGCTCTGCAAAATGATGCCGGTGCAGGTAAATG	1790
Qy	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle	380
Db	1791	ARGGGCTGAACAAATCCAAAGTTGATGAAGAAACAGCTGAATCATGTGGAAGCAATCAT	1850
Qy	381	GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys	400
Db	1851	AAATCAATGACTGTTCTTGAAGAAGACAGCCGGATATATCAATGACCGCCGGGAAG	1910
Qy	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
Db	1911	CGGATTGCAAAAGGAAGCGGACATCCGTACAGGAAGTCAACCGTCTGCTTAAGCAGTTT	1970
Qy	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys	440
Db	1971	GATGAATGAAAAAATGATGAAGACAGATGACAAACATGTCACAAAGGCAAGAAA	2024
Qy	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Db	2025	-----AAAGGTTTAAGCTACCTTTT	2045
RESULT 11	BSUB0009	BSUB0009	
LOCUS	Bacillus subtilis complete genome (section 9 of 21): from 1598421		
DEFINITION	to 1807200.		
ACCESSION	Z99112 AL009126		
VERSION	Z99112.1		
KEYWORDS	299112.1 GI:2633902		
SOURCE	Bacillus subtilis.		
ORGANISM	Bacillus subtilis		
REFERENCE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
AUTHORS	1 (bases 1 to 208780)		
	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,		
	Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,		
	Borriss, R., Bourisier, L., Brans, A., Braun, M., Brignell, S.C.,		
	Bron, S., Brouillet, S., Bruchi, C.V., Caldwell, B., Capuano, V.,		
	Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,		
	Daniel, R.A., Deniset, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,		
	Emmerston, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,		
	Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,		
	Galleron, N., Gim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,		
	Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,		
	Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F.,		
	Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,		
	Klaer-Bianchard, M., Klein, C., Kobayashi, Y., Koetter, P.,		
	Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,		
	Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,		
	Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P.,		
	Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,		
	Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,		
	Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,		
	Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,		
	Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,		

	Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,		
	Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,		
	Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,		
	Takeuchi, M., Tanakoshi, A., Tanaka, T., Terstra, P., Toponi, A.,		
	Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,		
	Viari, A., Wambit, R., Wedler, E., Wedler, H., Weitzenegger, T.,		
	Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,		
	Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and		
	Danchin, A.		
	The complete genome sequence of the gram-positive bacterium		
	Bacillus subtilis		
	Nature 390 (6657), 249-256 (1997)		
	98044033		
	9384377		
	2 (bases 1 to 208780)		
	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.		
	Direct Submission		
	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,		
	Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724		
	Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,		
	adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45		
	68 89 48		
	Location/Qualifiers		
	1. 208780		
	/organism="Bacillus subtilis"		
	/strain="168"		
	/db_xref="taxon:1423"		
	258. .4588		
	/gene="bpr"		
	258. 4559		
	/gene="bpr"		
	/EC_number="3.4.21.-"		
	/note="alternate gene name: bpf"		
	/codon_start=1		
	/transl_table=11		
	/product="bacillopeptidase F"		
	/protein_id="Csb13404.1"		
	/db_xref="GI:2633903"		
	/translation="MRKKTNRLLSSVLTWISSLFFGAAGASKVTSPVKKELO		
	SAESIONKISSLKSKFKKTKTLFKFDLANPKAAKAAKAAKSKLSAAKTEY		
	QKRSAYSSIKVTADESQQDVLKYLTKQKGNADQIHYYVNVNGIAHAKFVWKY		
	VQFEVRLVLPNEKOLFSSPMMKKAOKAIKATDGVENVNDQIDAPKAWALYDG		
	TGTVASIDTGVWENHPALKKRYGYPENPEFENEMWYDAVAGEWLPADLAHGT		
	HVTGTVGSEPDQNIQIYAPKAWIAKAFSEDGDTADILEAGEWLAPKDGNP		
	HPENADPVNNSGGSGLDWYRDWYNAWRAADIPFESAGNTDLFIPGPGSIANGP		
	ANYPEFATGATDINKLADFSLOGPSPIDEIKPEISAPGNIRSSVPQYEDGWDG		
	TSMAGPHVSAAALLQANASLSDVEMEDILTSTAFELTDSTFDPSPNGYGHGLVA		
	PDVASVTDGLKAGQYVSVDGDEPPYQHEKYTEAYEGGSLFLTLTADNVSVTS		
	VKLSKLDQGEWTEITAKRISGDHLKGTIQAETPDIKPKLSIKWMIHDFGVVSSD		
	YVDYVTKPSITAGYKODFETAPGGWASGNTNNWGVSPSTGPTAASGEKVTGNT		
	GNVANANLNMPPIKAPDSGSLFQFKSWHNLEDQDYGYYFVLPBGEKNWQAGV		
	YNGTSSWDEEIDLSAYKGNIOVMFNLSQDESIAKGGWIDDVVLSDKSGATYKK		
	NKLGVEPKSQKKKVPNPKAKPSANTAVKHONKALOPQLPLKAOVSUVTGKTYK		
	SDASTCGYILKHAGDYILMEARGYQSKTKYSLKTDQTQANTLBEAKKGLKGT		
	VINTGEPYTGASVYVDEARVAPAMNDKGEIMLEYEGAITKVAAPGYTSEDES		
	VELKRDVYKETALKPFVYGPGEIAYDDGTANANSYFAAGNWAYMTLADGDKGNL		
	TGGLFRWDTDFDQGEFFKVDYTDGDKGAPKKIAGPENALRWGWTGVDLS		
	SKGLMWDFLYLVYIQSKPDYSPGLAMDETQNSGRNWOYIDGWQPGKADGNMI		
	RALVDYAAVPEITPTSTDKSYTNKDSVTYVKNASPGTTVHYIYNGEKEAGPKAADGT		
	FHAGILIGENELATATSTONGTADASSPTVYTLQEKPELTLDNPKDGGKINKEL		
	TYGKASDDNLKDKVKNGKATVADGYSKILLLEGNREIKVIATDLAGNKTATKTV		
	IDVNFKPVISLIGPEKDLKAGESVKIAFSAEDLDTFTIRPLTNARASVQKAT		
	ELPLREISFGRYEGYWTATSSIKAKGKARVEIVRDDYDGNTEKTRKANGKNNTEN"		
	4571. .4588		
	/gene="bpr"		
	4754. 5683		
	/gene="spolIGA"		
	4754. 5683		
	/gene="spolIGA"		
	/function="processing of pro-sigma-E (SpoIIG) to active		
	sigma-E (stage II sporulation)"		
	/codon_start=1		
	terminator		
	gene		
	CDS		

```

/transl_table=11
/product="protease"
/protein_id="CAB13405.1"
/db_xref="GI:2633904"
/db_xref="SWISS-PROT:P13801"
/translation="MKTYLDVTLNFCFDALLLLTAFLTKRHKRRLVGGAFIQS
STVLLMFPSPSPIVEHPAGKLAFSVIVVTFGPKRPFPPFONLSEFYFAVFLANGGQI
IGASISIGSINQVGMATNQTGCDPISWLFYVGGFPALNPFPSKRRIEDIELEIKNQI
YERYSVQADLGSGTLHVRLGLDSGNQYLDLTPVPMIYIDKLEPIFGVAFETNIIR
NDPUEALEQDUSFRFLDKMKRLIPYGVGQGNQFLCVKPKDQHVTTMTKEEMISADXC
LIGSTYKLSADGEFDAILHPKMLSKRAKVS"
gene      5746..6465
CDS       5746..6465
/gene="sigP"
/gene="sigE"
/feature="early mother cell-specific gene expression"
/note="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKKILRLTHLWYKLLMKLGKSDVEVYIGGSEALPPPLSKDDEE
OVLMLKPGQQAARAILERILRLVWYIARKFENTINIEDLISIGTIGLIKAVNPF
NPEKXIKLATYASCTENETLMVLRNNKIRSEVSYEPLNDMDGNELLISDVLGPD
DOLIKDTEANVDKLLKALQELNREKQIMELRFLGVGEEKQKDVADMMGISQS
YISRLKRIIRLRKEFNKMV"
gene      6605..7387
CDS       6605..7387
/gene="sigG"
/gene="sigC"
/feature="late forespore-specific gene expression"
/note="alternate gene name: spoIIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSNRKVEICGVDTSKLPVLKNEMRKLFLQLODEGDSAREKLV
NGNLLSVLQRENNRGEYVDLFGVCGIGLMKSIDNFDLSHNKFSYAPVPIIGE
IRRYLRNNRPSRSLDIAYKALQVRELTISSETPAEIDIAKLVLEPHEIVFA
LDAIODPSYLPFIYNDGGDPIYVMDQISDERNTDSQWIBELALKEGMRLENDREKMI
LKRFFQCKTOMEVAEEIGISQAQYSRLEKAIQMKNNHQ"
gene      7335..8496
CDS       7335..8329
/gene="yima"
/gene="yIma"
/feature="unknown"
/note="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPTREMBL:O31723"
/translation="MILQIDNYSLRKNGKWLKIDHNKVEKENVYVLGNAGKTL
LNMKCSYFPTSGEMRYLGHFKTELGKLRKXIGLVSAALQKLYPADSAFIALS
GAYASIGYFPTSPKREKALGLEDGAIEXADRRYETLSQGEKORALLARALMADP
ELLIDPVTGKMDIAREKLLDFTIYANKENAPSILYVTHHAEELVFPVKALLLQK
GEVFGSGIKEMLDTQILSAFDTPIHVLWNQDRPFFVTRAEPITNA"
terminator 8472..8496
gene      8531..9811
CDS       8531..9811
/gene="yImb"
/gene="yImb"
/feature="unknown"
/note="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11

```

```

/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPTREMBL:O31724"
/translation="MDQIYSLQKVEEHEKEELIQLAKTLISYOTPPAPPARTNEGIOS
WIAVNLNGLFSIDKWDVPGDPNVVGRKLGTSADYSLIINGHVADVAVKDEEWK
HDPHPTEKNGLLIGRASDMKGMACVLFVAKLIREASIELPGDLILQSVIGEEVK
AGTLECCCKRCYHADRAIVADPISDMHIOGGGVITGWIEFKSOTPHGTRENHAGG
GTGCASAIEMKAKIAGLQGLERLHWSIMKSYFGFPGTNTNPVAVIGSGHRAFIADP
CRMVITVHFYFNKEDIVNLRNSDSINWLRNRNPFVFKWGGSSMIEDRGEIF
PALEVDHPGFVALTASHQVKECPIDVQSQVTDGGLYDAGIFCVIYGFGLHN
ABSYNKSVSTEQIOLVEYTRKILIDFTIISWCSRKEQ"
gene      9894..10139
CDS       9894..10139
/gene="yImc"
/gene="yImc"
/feature="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPTREMBL:O31725"
/translation="NLSISEFQVXDVNVNNGKKGISGIDIDINVTGKIQAIIIGGN"

Alignment Scores:
Pred. No.:      2,38e-89      Length:      208780
Score:          1647.00      Matches:      319
Percent Similarity: 84.62%      Conservative: 66
Best Local Similarity: 70.11%      Mismatches:   60
Query Match:     72.43%      Indels:       10
DB:              1           Gaps:          1

US-09-943-108a-2 (1-455) x BSUB00009 (1-208780)

QY   1  MetAlaPheGluClyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db   73153 ATGGCATTTTGAAGGATTAGCGACGAGCTGCACAGACGATTTCCTAAATTCGCGGAAAA 73212

QY   21  GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db   73213 GGGAAAGTCAGCGACACAGATGTAAAGACATGATCGCGTAGGTCCTGCTTGGCTGCTT 73272

QY   41  GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db   73273 GAGCTGACGTTAACTTTAAAGTAGTCAGGATTGTCTCAAAAAAGTAAGTAACGCGCT 73332

QY   61  LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db   73333 GTAGGCGAAGACGCTCATGAAAAGTCTGACGCGCGCGACAGGTCATTAAGTGTCTCAA 73392

QY   81  AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db   73393 GAGGAAGTCAGCTGAGCTGATGGCGGCGAAGAGACAAAATCGCGCGTCGCAAAAAGCGCG 73452

QY   101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db   73453 CGGACTGTATTATGATGGTCGCTCCAAAGTGGCGGTAAACACACACACACGCGGTAA 73512

QY   121 LeuAlaLeuLeuMetArgLysLysTyraAsnLysLysProMetLeuValAlaAlaAspIle 140
Db   73513 CTTCGGAATCTGCTCGGCAAAAACATAATCGCAACCGATGCTGGTGGTCCGCGATATT 73572

QY   141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db   73573 TACGCGCCAGCGCAATTAAGCAGCTGGAAAACACTCGGCAACAGCTTGATATGCTGCTGT 73632

QY   161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db   73633 TTCTCTCTTTGGCGATCAGTCTCTCTAGAAATAGCTAAACAGGCTATTGAGAAAGCC 73692

QY   181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db   73693 AAGGAAGACATATGATACGTATTGGATACGCGAGGCGCTTGCATATCGACCAT 73752

```

QY	201	AlaLeuMetAsnGlnLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu	220
		:	
Db	73753	GAACGTGATGATGAATCAGCAACGTCACAAAGTAATCCGAATCCGGAAGAAATTTCTCGT	73812
		:	
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
		:	
Db	73813	GTTGTCGATTCATGACCGGTGAGACGCTGTGAATGTTGCCAAAGCTTTAATGAACAG	73872
		:	
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla	260
		:	
Db	73873	CTCCGTTTAAACCGGTGTTGTGTGATAGCTGATGGAGACACACAGCGCGGGCGGTGG	73932
		:	
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
		:	
Db	73933	CTTTCTATTCCGCGTGTCAACACAGCCCAATTAAGTTTGCAGGTTTGGCGGAAAAGCTT	73992
		:	
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
		:	
Db	73993	GATGCGGTAGAGCCGTTCCATCTCTGAACGATGCGATCAAGGATTCCTCGCATGGCGAC	74052
		:	
QY	301	ValLeuSerLeuIleGlnLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu	320
		:	
Db	74053	GTGCTGCATTTGATTGAATAAAGCAGCGCAGCGTGTATGAAGAACCAAGAGAGCTG	74112
		:	
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
		:	
Db	74113	GAACAAAAAATGAGAACGATGAGCTTCACATTTGACAGATTTCTGAGCAGCTCGGCGAA	74172
		:	
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
		:	
Db	74173	GTCAAGAACATGGGGCGGCTTGATGAGCTTCTCAATGATGTCGGGTGAGGTAAAGATG	74232
		:	
QY	361	LysGlyLeuAspLysLysMetSerGlnLysGlnIleAspHisIleLysAlaIleIle	380
		:	
Db	74233	TAAGGCGCTGAATAACATCCAGTTGTGTAATAACAGCTGAATCATGTGGAAGCAATCATC	74292
		:	
QY	381	GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys	400
		:	
Db	74293	AAATCAATGACTGTCTTGAAAAGAACAGCGGATATTCAATCAATCCAGCGCGCGGAG	74352
		:	
QY	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
		:	
Db	74353	CGATGTCAAAAGGAGCGGGACATCCGTACAGGAAGTCAACCGTCTGCTTAAGCAGTTT	74412
		:	
QY	421	AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys	440
		:	
Db	74413	GATGAATGAATAAAATGATGAAGCAGATGACAAACATGTCAAAAGGCAAGAAA-----	74466
		:	
QY	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
		:	
Db	74467	-----AAAGGGTTTAAGCTACCTTTT 74487	
		:	
RESULT	12		
LOCUS	AP001515/c	303249 bp	DNA linear BCT 01-FEB-2001
DEFINITION	Bacillus halodurans genomic DNA, section 9/14.		
ACCESSION	AP001515 BA000004		
VERSION	AP001515.2		
KEYWORDS	GI:12641879		
SOURCE			
ORGANISM	Bacillus halodurans DNA.		
REFERENCE	Bacillus halodurans		
AUTHORS	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
TITLE	1 (sites)		
JOURNAL	Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fujii,F.,		
MEDLINE	Nakamura,Y. and Inoue,A.		
PUBMED	An improved physical and genetic map of the genome of alkaliphilic		
REFERENCE	Bacillus sp. C-125		
	Extremophiles 3 (1), 21-28 (1999)		
	99184645		
	10086941		
	2 (sites)		
	Takami,H. and Horikoshi,K.		

FEATURES
sourceLocation/Qualifiers
1. .303249/organism="Bacillus halodurans"
/db_xref="taxon.86665"

/note="alkaliphile"

104. .1102

/gene="BH2267"

104. .1102

/gene="BH2267"

/codon_start=1

/transl_table=11

/product="penicillin-binding protein"

/protein_id="BAB05986.1"

/db_xref="GI:10174887"

/translation="MKNHLHTIMEDWKLGSVALMKKGEDIPPIASLGANRAERIPNE
HTRFGIAGCKLFTALICQVVEAGKLSFTPLSLDWLADPPNVTIHLHLTHSGYP
DYFDEITDDFEDLWQDVWHLRLKDFLFLGHAPKPKFPGHFRPHNNAQTILLGL
VYESVGVTQEVVEANVFORAGMHESGIFAFDLTPAKTALGYIDLEQSGKTIWLSL
PVIQSDGGAYVAEDMMKLWALMRHELLNETYTKLLTPPHVCEDDDYGYGWNK
QODGAIKSYHMVGYDFGVCFHSAFYFTSNGIVVVCANQSSGAYDVMAAIEALFSEA"
complement(1146. .1469)

/gene="BH2268"

complement(1146. .1469)

/gene="BH2268"

/note="BH2268"

unknown

/codon_start=1

/transl_table=11

/protein_id="BAB05987.1"

/db_xref="GI:10174888"

/translation="MKVITLCGSTKFKRQFQASAYLTLLKGHVLSVGFQSGIET
TEEQVQMFENIHRKRIEMSGDEIFVLVDVGYIGRSTRKEIFATSKGMPRIYSESLA
STTFS"

complement(1636. .2373)

/gene="BH2269"

complement(1636. .2373)

/gene="BH2269"

/note="BH2269"

unknown

/codon_start=1

/transl_table=11

/protein_id="BAB05988.1"

/db_xref="GI:10174889"

/translation="MYKKVILGVFALSHGLMACNQSQEQEEVEKNIDGEASGIPSI
BEIDINLDPVATVNGESIDKETVMVLFQOMTMAQVGI:SLGEEGKQMSLIFE
QTIDQLNLOQAAATEKEIEVSEEDRELATFVAOFSSEALMALEAGSGSNDP
LREIEHVHQKYVEETEAINTVEEIQARVEEKQYITDELPTEEQLQSNIEOR
LIAKKEQEQLEVEKLRDEGDIIVHI"

complement(2582. .3367)

/gene="BH2270"

complement(2582. .3367)

/gene="BH2270"

/codon_start=1

/transl_table=11

/product="ABC transporter (ATP-binding protein)"

/protein_id="BAB05989.1"

/db_xref="GI:10174890"

/translation="MADILQVQTVLRKAKKPILOQISMQVQEQGBHWAIVGLNSGKT
SLKLIITGYEWPTEGHVAVNKRIGHVPIDQVRKRIGWYSMSLDDRPFTTKGQDTVLEI
VLSGLHGTGVYPLGEADVAKARAETETGLHLDERELFALSQGEKRAFARAVY
AKPEFLIDPFTTGLDLIAREQLLOTLESVATKGAFTLYVTRIFPEIQFIITHVM
LKDGHVTAQGRKEDEVLTKEMITAFGCSLDVIAQSGRIWVKPM"

complement(3378. .4310)

/gene="BH2271"

complement(3378. .4310)

/gene="BH2271"

/codon_start=1

/transl_table=11

/product="dipeptidase"

/protein_id="BAB05990.1"

/db_xref="GI:10174891"

/translation="MEIFDAHCDALLKLWEDPAREYTDPAIEANAVRLMGGVRAQC
FAIPVGNIFDQKQFALQQAELFHEKILKKHPQPSFIERWDDLLTLDDAQIGAILT

gene

CDS

gene

CDS

LEGADPFDDPTKLTLYRLGYRSLGLTWNQANLCADGVGCEPRGGGLTEWGAUVSLN
NEHVLTDSVHLSERGFWDVIEYAQFPAVPHSNKAICKHRNLTCAOLKALIEAGCF
IGIYVHPHFLLTCTDDASIDDYMKHIEHCISLGQKEVVGSGDFDGIPICFVGLSHAGR
FCWFMDLARIYDEELVRGFAWDFLKATPLSYV"

4406. .4840

/gene="BH2272"

4406. .4840

/gene="BH2272"

/codon_start=1

/transl_table=11

/product="heat shock protein class I (low molecular

weight)"

/protein_id="BAB05991.1"

/db_xref="GI:10174892"

/translation="MSREESGSKLPNVYQDFLRSDDDFSKSFQHLHDNMGFAPSTP
IRQKEKGEQIIIEAELPGVSNQIQLDVFFQYVIRIHEEVELEKREDEKGIYSKSL
SQAEKRVIPFPYVQESDIKARLNRGLLTIAIPNHRKRINIE"

4864. .5193

/gene="BH2273"

4864. .5193

/gene="BH2273"

/note="BH2273"

unknown

/codon_start=1

/transl_table=11

/protein_id="BAB05992.1"

/db_xref="GI:10174893"

/translation="MYNLHPQTVSQGAYVQQPFQHSHPNHPQPTSSGQWMTSTG
FSFPFPNPFYSQQKRPASPKAAFTNQDGTDFINKTFQTVQVVKTVNQVPLVKS
VGSFIR"

complement(5202. .5510)

/gene="BH2274"

complement(5202. .5510)

/gene="BH2274"

/note="BH2274"

unknown

Alignment Scores:

Pred. No.: 2,43e-85 Length: 303249

Score: 1584.00 Matches: 307

Percent Similarity: 83.33% Conservative: 73

Best Local Similarity: 67.32% Mismatches: 70

Query Match: 69.66% Indels: 6

DB: 1 Gaps: 2

US-09-943-108a-2 (1-455) x AP001515 (1-303249)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20

Db 210434 ATGCGCTTTGAAAGTTTAGCGGAGCGCTGCAAGATACGTTGAAGAAAATTCGCGGAAA 210375

Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

Db 210374 GGAAGTAGTAGTGAACAGACGTAAAGAAATGATCGGTGAAGTTCGCTGCCTCCTTA 210315

Qy 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60

Db 210314 GAACGGATGTTAAACTTTAAAGTGGTGAACCAATCATCGCAATGTGAAGAAAAGCC 210255

Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80

Db 210254 CTCGGTCAGAAAGTCATGAAAGCTTGACACCTGGACAACAGGTCATCAAGTCGTCAT 210195

Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100

Db 210194 GAGAGCTGACGGCGCTGATGGGAGCGGAGCAAAAGCAAAATTCGGTTCGACCAAGCCG 210135

Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120

Db 210134 CCAACGGTCGTGATGTCGTCTTCAAGGGGCTGGGAAAACGACGACGCGGCGAAA 210075

Qy 121 LeuAlaLeuMetMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

||||| ::::::::::::::::::::: ||::::::::::|||

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaLea 260
 Db 721 CTGGTGTGACAGGTGTCTATCTTACGAAGCTTGACGCGATACTAGAGGGGAGCGGCC 780
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
 Db 781 TTGTCAATCGCGCGCTGACAAACACGCGATCAAAATCCCGGAATGGCGAAAGCTT 840
 QY 281 AspGlyLeuLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 GATCGGTGAGCGCGTTTCACCTTGACGAGATGCTTCCGGATCTCGGAATGGCGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 GTGCTCACCTTATCGAGAAAGCCGACGCCAAGCTAGACGAAGAAAGCGAAGCACTC 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLysAspAspPheLeuGluGlnLeuAspGln 340
 Db 961 GAGCAAAAATGAAATAATATGAGCTTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTCCGCATATGCGCGCGCTTGAGACACTGATCAATAATGATCCGCGCGAGCGAAATG 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAla 378
 Db 1081 AAAGGCTTAAAAAAGCTGAAGTTGAGTAAAAACAGCTCAGCCATATCGAAGCG 1134
 RESULT 14
 AB007684
 LOCUS
 DEFINITION Clostridium acetobutylicum ATCC824 section 172 of 356 of the complete genome.
 ACCESSION AB007684
 VERSION AB007684.1
 KEYWORDS GI:15024715
 SOURCE
 ORGANISM Clostridium acetobutylicum.
 Clostridium acetobutylicum
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
 REFERENCE
 1 (bases 1 to 11024)
 AUTHORS Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I., Tatusov,R.D., Sabathe,F., Doucette-Stamm,L., Soucaille,P., Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
 TITLE Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum
 JOURNAL J. Bacteriol. 183 (16), 4823-4838 (2001)
 MEDLINE 21359325
 PUBMED 11466286
 REFERENCE
 2 (bases 1 to 11024)
 AUTHORS Childress,D., Zeng,Q. and Smith,D.R.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA
 FEATURES
 source
 1..11024
 /organism="Clostridium acetobutylicum"
 /strain="ATCC 824"
 /db_xref="ATCC:824"
 /db_xref="taxon:1488"
 84..317
 /gene="CAC1747"
 /note="acpA"
 84..317
 /gene="CAC1747"
 /codon_start=1
 /transl_table=11
 /product="Acyl carrier protein, ACP"
 /protein_id="AAK79713.1"
 /db_xref="GI:15024716"
 /translation="MWFEVKDIIADQLGIDATEIKMESSFDDIGADSLIVELIMA

LEEFEDTETPDEEAERKVSVDGVVYKKAHTEE"
 444..1136
 /gene="CAC1748"
 /note="rncS"
 444..1136
 /gene="CAC1748"
 /codon_start=1
 /transl_table=11
 /product="DsrRNA-specific ribonuclease III"
 /protein_id="AAK79714.1"
 /db_xref="GI:15024717"
 /translation="MEERLLRELESKLGIEFQINLLVTAHTSSYANENKNAEYNE
 RLEFIDAVLQISITVSFFPKYPTTISEGELTKKRALVCGMSLHISGERWOLGVISM
 SHGELNGRTRYSIIADCEAVIAIYLDKGFDTAKNPILREFEGTIONAEVKIIL
 DYKRLQELLOSKGKTDIKYTLVRHEGPPHRRKFFVNLNPDNDVASTGEGITTKDAEQ
 DAACKALGGLDN"
 1137..2213
 /gene="CAC1749"
 1137..2213
 /gene="CAC1749"
 /codon_start=1
 /transl_table=11
 /product="BioB/LipA-like protein"
 /protein_id="AAK79715.1"
 /db_xref="GI:15024718"
 /translation="MGKSHVITIPVRHEGCPHPCVNCQNTITGEESKVNQGVKST
 VEDLDIIPNKSATIEISFGGTFTAINIQKELLISVARHYKEIGKIKYIRLSTFRD
 YIDDEILNLKNYSVDIIELGVQSDSEVLKSGRHIEDVYKASRLIKYIGTLGH
 QIMGLPQNAEKDIEITRKVICMKNPCIRIYPALVIKDTAMEKMNIGIYKPYDIBE
 CVELSKVFMGLFANDIEVIRIGLOPTEINTDELGVAGFPFAPRELVEGSIANNEM
 FDNLKCESNTVIDYLNPKDIDISKLYSNKKTFFYSMKDKLNLRLINVIQNPEVENDILK
 NNFNDCKIYSKYEYLKVKYKEGK"
 2270..2530
 /gene="CAC1750"
 2270..2530
 /gene="CAC1750"
 /codon_start=1
 /transl_table=11
 /product="SpoVS-related protein"
 /protein_id="AAK79716.1"
 /db_xref="GI:15024719"
 /translation="MEVLKVSQKSPKSVAGALAAVLRLNSAEVQAVGAGAINQAYK
 AIAITRGVAPNGIDLVVVPAPAEISIEGDRTAIRKPIVESR"
 2747..6322
 /gene="CAC1751"
 /note="smc"
 2747..6322
 /gene="CAC1751"
 /codon_start=1
 /transl_table=11
 /product="Chromosome segregation SMC protein, ATPase"
 /protein_id="AAK79717.1"
 /db_xref="GI:15024720"
 /translation="MFLKSIETRGKSPADKIDILFKNGITAVVGPNSGKSNISDAV
 LWLIGSVNLRGGMEDVIFAGQYKRSVQAQVSLIDNSDKQLMDLSEYTVSR
 RLYRSGSEYIINNTKRLDQELFGDTGIGKEGYSIIQGGKIEAVISGPERRAL
 LEEAGIVAFKTRVDAEKLENTQNLVINDILITRERLEPRLESRAKAFVSL
 SDELKTEINTIISIDNIDYNDLQKMAIDLKSIDENVKDKKISLELKVATBSL
 DEFDAKYSNKTIVYESKSEHOKILSRTELLAKETSNQVAKNKLKRIEOLDNSIVN
 LKSYEIQLKITLEDKNYNKELLSKINKSEKKKNIDGLLEWEKSIKQYKDAIDIL
 STISQNNNEVILKIEISNESKSLIKRAGEYSKSLINEVTRNTISELVIDNKL
 ISGYENQITRNRKISKLRIISDEKLNRELNSKNLKNLANKNMLINLEKQYVNR
 SVNLGMHVTGFDVDPSSFLGEVTKKFEFATVETSLGAAISDITIDDNIAK
 KLINLAKSNLGRATPLNIIKGRKLNISDATHREKFGIASELIDYDTFTLPVYN
 VVLGRTVIYDNMDSALKIAKNSYSFKIVITLGEVNVPGSLTGGSTYSKASLIIGK
 REBELNELNYSQALEOSSNKKIENKVKVKEDNLCDLDTDTTHGKIELTKIKER
 LKSIDISEKLNKSYINTSVGEIGFKERINKHLEKLVKEEENKALKUREANNLNLD
 ELERKLKDNSKVINNEISMKYDKAKSDEMISMSTREIERYKVENHNNENKISL
 KNEIDDFENQVKILNKKIYNNQVQIQVQIPELNSKQSEVSKSEKNIENKEN
 SLOGINVLQAKETEKRHYENIAKITESENLYERLNSDPKLTYSRAEFKEIIEEM
 ISYKKKIDELKREINMGVNVVASTEEYKVKKYTFMNSQKEDLDNKADELIVIBE
 MTSKMRIVFENENFKLNFKITFDKULFGKGNADLILSGDDEINSSIEINPEPKKL
 QNINLMGSGKGKLSAALLFAILKMKPTTFCILDETEAALDADNARYAEFLRFSFN


```
MLDAFGVKAHEVKIIDVTNQDIREQLGIDTSKPIPASSQSISISSVYVVKDKGINVFT
NNLEFVSGMIANALTSGVNADVKADPKVTGTAALAGILKGFEDASGEELSIPK
KEAREEISITNNLSNAKTQOGLIKDEAAVVNDIKTDVILKDKPKNDEBIGLVNN
VIRNINILFQGGQEGTIFKFSKINDIDINAGKESLQMDKILQILKIDTIGKQLEE
SGLLEKALNGIKNIVDIDKFLVNNFSSASEKVKDGTIDENGIVIAITGNSDESKN
BESIQDPQTQSNNDNNOENPOGQNTK"
complement(4141..7578)
/gene="spovK"
CDS
complement(4141..7578)
/gene="spovK"
/note="1145 aa, similar to gp:AB015670.7 Bacillus sp.
genes for Cbase, Cgrase, MBP and 15 ORFs, partial and
complete cds from Bacillus sp (549 aa); 45% identity in
522 aa overlap Also similar to prf:2516401X stage V
sporulation protein K from Bacillus cereus
CPE1547"
/codon_start=1
/transl_table=11
/product="stage V sporulation protein K"
/protein_id="BAB81253.1"
/db_xref="GI:18145210"
/translation="MLQKNIFVERKDMVKNKEIENGFFKIEKKLNEETIGOKTFNKL
CDYFKIEDSGKILLIIVEREDIFKNSVKYFPEELNYKFKVNSKIDEIDLAAYF
NLGNAFLTDLYELKNDSCQFLFKNTDKASEDILNLVSGISPNSCIMLKNQYIKNK
FLVAEBIDEDIVNKKIICNNKFFLYKNDKGBELEIDNEFLINKDKTLRKALSEK
DNKVIKKLLKINKVKELFDIDIIIGVDENPDVREKSGVCTYIEDNFRKGSLSLD
QYISKICKPILMIIRKEALEANSRLIYVEDDIFCKSDNDEVYDLSKYSPLEEVYR
YKLSIIGMKLEIDKLENNFKVKIRKGLKTKTQISLNMIFAGNAGTKNAAR
ITFDYVALGNIKENIFKEVSKADIFIGESINTARTNETIIESALGGVLFIDEAYSIC
TDENKAGRIEDVALLKIEDNRNDITVILAGYEKDMETFLSQGLKSRPPIKFE
DYNSEMETAVNIAKSGYRIAKSAKAGLIEFLTKNQMGVSKDGNARFVNIVENA
IWDASKYLVNKRATDILLEKDNFNFKVSAKFDLEKLEIIGLEPVEKFEFRSOKLI
VAQKRKSGVGNPKIRQNLNMFAGNPGTGKYSIARLYAEMLSGILLKVGQLYETDR
SVFSDIPGTESKRTEKFEKALGGVLFIDEAYTANDSISGREALITLLEKEDISRE
VIVILAGEEMENFTEDYVIGLSRFPWTKEFDYNPELLEMAIKLVEAKGLSKN
IKANNAIKYDNDYHEDAKQSGRMVRNRYVQLIRNQSIRIAESDISVYEMNLNKD
NMVFTSGSGKTKVNTLYSEMLYSGILIAKKNIVQDKYEPAMINSCKVEDILANK
HYGLIYIDKKNFTGEERYNEIVSDLIKFDNNSNRIVIVLGGIRGKMKDILITNEG
LNYRFDVDFGNYNERYDIASSTLVKKGFTIDEEAETALENAIGDIQKRWGELAL
KNGMLQFLDSLYRVQSRVCEDFDINTNRIISEDILKSTEIFLKNNTTQNGSF"
complement(7634..9622)
/gene="fdh"
complement(7634..9622)
/gene="fdh"
/note="662 aa, similar to pir:S66017 formate dehydrogenase
homolog YvyaE from Bacillus subtilis (667 aa); 32.1%
identity in 667 aa overlap
CPE1548"
/codon_start=1
/transl_table=11
/product="probable formate dehydrogenase"
/protein_id="BAB81254.1"
/db_xref="GI:18145211"
/translation="MEVLSHGCTDLCPCCKFNVYKEGSEILKIEGDKBHPPTKGLIC
KKGVAHLNRLNKHDRIVTPLLKNGVWEIEIFEDALIMKEKLEYTKVYSKSLTLY
SOYSGGVKLGIEDIFENFYGGSVKATGCPGCPWAGMRAQKIDFGDSVSNSEDMNSK
NIFLMGKNPANTIIHTMALINKAKNGSKLIVDPINTQSAKLGDIHVAIKFEGTAL
AMAKAIISKGLQKDFINKYVLFQFQYKQHLNFDLVDLSECGIEIEDIEKLTLY
YCKRNSIYLGGMOKYKNGGNTIRADALGALTQIGVAKGGVYANKVLSRILSDS
PKSGEVENREFYVSNINIEDEPKYSLSVEDSNAPIKIMVIANSNLNMQLPNLR
LANSIDKYERKVCDFMDTASCKDLPTPCTNLTESDGMVFSMTNPNLYIYNKIIIE
PREKLDIYFFRELAKRMMLKGYPSLSKDYLSKVIEPLMSYKNDITLDYLNKPPFT
VYDDVAWENKFKTPSGFELAKRALKECGLPTLYLSPIRENCFRLLTHNSKDSL
SSQHYIDVEKAKYVLNENMRKRYFLICGERVKLSKRTGRTAICSLDNGVQDYVALM
YVGMWKKGNPNFLTESGISDMGGQITYNTEPIETEN"
10033..11127
/gene="prs"
CDS
10033..11127
/gene="prs"
/note="364 aa, similar to pir:A71345 probable
phosphoribosyl pyrophosphate synthetase (prs) from
Treponema pallidum (421 aa); 32.3% identity in 375 aa
```

```
overlap
CPE1549"
/codon_start=1
/transl_table=11
/product="probable phosphoribosyl pyrophosphate
synthetase"
/protein_id="BAB81255.1"
/db_xref="GI:18145212"
/translation="MNEQNHGILIALESCTELGNAIDKLIQKRNCESEFLIKTDEI
RFSNGEKVKIPQSVRGRIYIILCDIGNYCTYKMFGENKHPDEHQDIKRPVQA
RKGARRITVIMPLIYESRQHRKRGESLDCALQLELRIGVDVLTDFDHPNVQNA
IPKSFENFYPTDYIVKSLIHNEKOLELDEKELIVTSPDTGAMDRAIYYSVLGVQVG
LIFYKRHSRITVNGKNPIVKHYEMGRDVGDKVLIVDMIASGESYLDIAKELGRNA
RNVYVATTFSEFTGELEKFNKFNFDGIKISVYSTNLYIPEELNAPWFKAVDLSDFI
ARIINRLNYDKSVASYMDATNLIQRLDLEK"
complement(11177..12238)
/gene="CPE1550"
CDS
complement(11177..12238)
/gene="CPE1550"
/note="353 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81256.1"
Alignment Scores:
Pred. No.: 4.39e-73 Length: 296750
Score: 1382.00 Matches: 279
Percent Similarity: 78.88% Conservative: 72
Best Local Similarity: 62.70% Mismatches: 86
Query Match: 60.77% Indels: 8
DB: 1 Gaps: 5
US-09-943-108a-2 (1-455) x AP003191 (1-296750)
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgLys 20
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 185046 ATGCGCTTTTGATGATTAGCGCTAAATATACAGACACGCGTTAAAAATTAAGGCTAAA 184987
QY 21 GlyLysLeuThrGluAlaAspLysLysLysLysLysLysLysLysLysLysLys 40
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184986 GGTAAAGCTAACAGAAAAAGATATAAAGACCCATGAGAGAAAGTAAAGCTTGCCTTTTA 184927
QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184926 GAGCAGCATGTTAACTTCAGGTGTTAAAGAGTTTATATCAAAATGTTAAAGATAAATGT 184867
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184866 GTTGGTGAAGGAGTATTAAATAGTTTAACTCCAGGTCAACAGGTTATTAAAAATAGTTAAT 184807
QY 81 AspGlnLeuThrLysLeuMetGlyGlyGlnAsnThrSerIleAsnMetSerAsnLysPro 100
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184806 GACGAATTAACAACCTCTTATGGGAGAGAGCTGAGAGTAATAATAACTCAGACAAATGGA 184747
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184746 CCACAGTGTTTCATGTTAGTGTAGTCTTCAAGCTCAGGTAAAAACACTATGCGTGGAAAA 184687
QY 121 LeuAlaLeuLeuMetArgLysLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184686 CTTGCAATTCACCTTAAGAAAAAGAAA---AACAAAAAGCCATTACTTGTGGCTTGTGATATA 184630
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184629 TATAGACTCGACGTAAAGCAATATACAGGTTGTAGAAAAACAATATATACCATTA 184570
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
Db 184569 TTCTCAATGGGAGATAAGGTAAAGGCTGTAGATATGCCAAAGCTGCATAGACATGCA 184510
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||...|||||...:|||||...:|||||...:|||||...:|||||...
```

Search completed: February 25, 2003, 02:59:00
Job time : 4087 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 00:19:11 ; Search time 296 Seconds
(without alignments)
3461.687 Million cell updates/sec

Title: US-09-943-108a-2
Perfect score: 2274
Sequence: 1 WAFEGLSERLQATWQKMRGK.....GKKGKRQNMKLGNLFF 455

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xip
-O=/cyn2_1/USPTO.spool/US09943108/Runat_21022003_151629_1790/app_query.fasta_1.647
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09943108.@CN_1.1.396.@runat_21022003_151629_1790 -NCPU=6 -ICPU=3
-NO_XLIFY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2274	100.0	1368	20	S. aureus fff codi
2	2261	99.4	1368	23	Staphylococcus aur
3	2235	98.3	1365	23	Staphylococcus aur
4	2235	98.3	1368	23	Staphylococcus aur
5	1719.5	75.6	2256	24	Listeria monocytog
6	1708.5	75.1	319630	24	Listeria innocua c
7	1534	67.5	1014	24	Staphylococcus epi
8	1418.5	62.4	1416	23	Enterococcus faeca
9	1418.5	62.4	1434	23	Enterococcus faeca
10	1417.5	62.3	6729	20	Bacillus lichenifo
11	1402	61.7	1136	24	2CFE 28 coding seq
12	1312	57.7	1605	22	Streptococcus pneu
13	1307	57.5	1572	23	Streptococcus pneu
14	1306	57.4	1395	23	Streptococcus pneu
15	1306	57.4	1572	22	CFE 28 coding sequ
16	1306	57.4	7577	19	Streptococcus pneu
17	1305	57.4	792	20	S. aureus fff codi
18	1297	57.0	1569	20	Streptococcus pneu
19	1297	57.0	1569	24	DNA encoding strep
20	1279.5	56.3	1563	24	Streptococcus poly
21	1273.5	56.0	1563	24	Streptococcus poly
22	1266.5	55.7	2026	19	S. pneumoniae deri
23	1266.5	55.7	2026	19	Streptococcus pneu
24	1168	51.4	1641	22	C glutamicum codin
25	1168	51.4	349980	22	C glutamicum codin
26	1164.5	51.2	1362	22	E. coli growth and
27	1164.5	51.2	1362	23	E. coli DNA for ce
28	1158	50.9	721	23	Staphylococcus aur
29	1150	50.6	1374	23	Pseudomonas aerugi
30	1138	50.0	1389	23	Haemophilus influe
31	1138	50.0	1830121	17	Haemophilus influe
32	1133.5	49.8	1506	23	Salmonella typhi D
33	1131	49.7	8367	21	N. meningitidis pa
34	1131	49.7	349980	21	Neisseria meningit
35	1131	49.7	349980	21	Neisseria meningit
36	1115	49.0	92407	22	Genomic fragment #
37	1014	44.6	21567	23	Propionibacterium
38	1007	44.3	644	18	Staphylococcus aur
39	1006	44.2	640681	24	Buchnera sp. genom
40	907	39.9	910715	20	Borrelia burgdorfe
41	890	39.1	580073	18	Mycoplasma genital
42	854	37.6	1347	23	Helicobacter pylor
43	853	37.5	4804	20	Enrichia sp. E74
44	847	37.2	521	23	Staphylococcus aur
45	847	37.2	521	23	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAAX27221
ID AAAX27221 standard; DNA; 1368 BP.
XX AAAX27221;
AC AAAX27221;
XX
DT 28-MAY-1999 (first entry)
XX
DE S. aureus fff coding sequence.
XX
KW Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;
KW immunological response; gene therapy; infection; otitis media;
KW conjunctivitis; toxic shock syndrome; septic arthritis; ss.
XX
OS Staphylococcus aureus.
XX
PN EP902087-A2.
XX

PD 17-MAR-1999.
 XX PF 24-AUG-1998; 98EP-0306741.
 XX PR 10-SEP-1997; 97US-0927216.
 XX (SMIK) SMITHKLINE BEECHAM.
 XX PA
 XX PI Wallis NG;
 XX WPI: 1999-169238/15.
 DR P-PSDB; AY00910.
 DR
 XX
 PT New Staphylococcus aureus Signal Recognition Particle (SRP) with
 PT protein (ffh) and RNA (ffs) components - the SRP gene and protein
 PT useful as diagnostic reagents and for prevention and treatment of
 PT Staphylococci infections which cause otitis media, septic arthritis
 PT and toxic shock syndrome
 XX
 PS Claim 1; Page 28; 35pp; English.
 PS
 XX
 CC This sequence encodes the staphylococcus aureus signal recognition
 CC particle (SRP) ffh component. Ffh polynucleotides are useful for
 CC diagnosing a disease related to expression of ffh polypeptides by
 CC analysing for the presence/amount of ffh protein in a sample due to
 CC infection of a micro-organism with the gene, or determining the nucleic
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful
 CC for treatment of an individual in need (polypeptide) of, or needing to
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and
 CC polynucleotides are useful for identifying agonists and antagonists by
 CC binding and observing the affect of ffh polypeptide activity, which are
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments
 CC are also useful for inducing an immunological response (T cell/antibody)
 CC to protect against disease, by direct administration (vaccine), or via a
 CC vector (gene therapy). Anti-fhh antibodies are useful as antagonists, and
 CC for protecting against disease. Diseases diagnosed, prevented and treated
 CC include those caused by infection, especially bacterial infection,
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and
 CC are useful for bathing wounds and implants prior to surgical
 CC implantation.
 XX
 SQ Sequence 1368 BP; 521 A; 186 C; 288 G; 373 T; 0 other;

Alignment Scores:

Pred. No.: 3 47e-197 Length: 1368
 Score: 2274.00 Matches: 455
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 20 Gaps: 0

US-09-943-108a-2 (1-455) x AX27221 (1-1368)

QY 1 MetalAlpheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCAATTGAGGCTATCAGACGCGCTGCAAGCGAGATGCAAAAATGGTGGTAAAG 60
 QY 21 GlyLysLeuThrGluAlaAspPheLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAAGATTACGCTATTT 120
 QY 41 GluAlaAspValAsnPhelLysValLysGluPheLysThrValSerGluArgAla 60
 Db 121 GAGGCTGACGTAACTTTAAAGTGGTAAAGAAATTTATTTAAACATATCAGACGGCGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLeuValGln 80
 Db 181 TTAGGTTCCGATGTAAATGCAATCATTAAACACCGGCAACAAGTTATTAAATACTTCAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100
 Db 241 GATGAATTAACGAGTGTGAGTGGAGAGAAATATCATCGATTATATGTCATAAATAACCA 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlnValAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCTACTGTGTGTAAGATGGTGTGTACAGGTGCTGTGTAAACAACTATGCGAGTAAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspPhe 140
 Db 361 TTAGCATTATTGATGCGTAAAAAATACAAACAAAACCTATGTTAGTTCACAGCATATT 420
 QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysPheProVal 160
 Db 421 TATGTCOCAGCAGCATATAATCAATACAAACAGTAGGGAACAAATTCATTCTCTGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
 Db 481 TACAGTGAAGAGATGATGAGTAAAGCCACAAATTTGTAATCATGATTAACCATGCT 540
 QY 181 LysGluGluHisLeuAspPheValLeuLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200
 Db 541 AAAGAAGAACATTAGACTTTGTGTAATCATGTATACAGCAGGTGCTGATACATCGATGAA 600
 QY 201 AlaLeuMetAsnGluLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220
 Db 601 GCATTGATGACGAATTAAGAAGTAAAGAAATTCGTAACCAACGAAATTTATGTTA 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 661 GTTGTCGATTCATGACGGGTCAAGATGCTGTCATGTCGAGATCTTTGACAGTCAA 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
 Db 721 CTTGATGTCACAGGTGTACCTTAACATAATAGATGCTGATACACGTGGTGGTGCAGCT 780
 QY 261 LeuSerLeuArgSerValThrLysProLysPheValGlyMetSerGluLysLeu 280
 Db 781 TTATCTATTCTGTCGTCACAAAAACCAATTAATTTGTTGTTGATGAGTGAAGAAGTTA 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLeuGlyMetGlyAsp 300
 Db 841 GATGTTTTAGAGCTATTTCATCTGACGTATGCGATCAGCTATTTTAGGTATGGGTGAT 900
 QY 301 ValLeuSerLeuLeuGlyLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
 Db 901 GGTGTTAAGTTTAAATGAAAAGCGCAACAGATGTGATCAAGAAAACCAAAAGATTTA 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
 Db 961 GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTTAGAACACACTTGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspPheLeuMetLysMetLysProGlyMetAsnLysMet 360
 Db 1021 GTGAAAAATCTAGGACCACTGGATGATATTTGAAAATGATTCACAGGTATGATTAATG 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLeuAspHisLeuLysAlaLeuLe 380
 Db 1081 AAAAGGCTAGATAAGCTTAATATAGTGAAAAGCAAAATTCATCATATTAAAGCGATTATC 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 CAGTCAATAGCGCGCTGAAAAGAAACCAATCCACACATGTAATGATCAGTAAAAAG 1200
 QY 401 AsgLeuAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATTCGTAAGGGTCTGGTGGTTCATTACAGAAGTCATCCTGTTGATGAACATATT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLys 440
 Db 1261 AACGATATGAAGAAAATGATGAACAAATTCACTGGTGGCGTAAAGGTAAAAAGGTAAA 1320
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAAATTTACCGTTT 1365

RESULT 2

AAS55331
ID AAS55331 standard; DNA; 1368 BP.

XX AC AAS55331;

XX DT 13-FEB-2002 (first entry)

XX DE Staphylococcus aureus DNA for cellular proliferation protein #1643.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI: 2001-611495/70.

XX DR P-PSDB; RAU37472.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 8968; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences.

XX Seq Sequence 1368 BP; 517 A; 187 C; 291 G; 373 T; 0 other;

Alignment Scores:

Pred. No.:	5,26e-196	Length:	1368
Score:	2261.00	Matches:	452
Percent Similarity:	99.78%	Conservative:	2
Best Local Similarity:	99.34%	Mismatches:	1
Query Match:	99.43%	Indels:	0
DB:	23	Gaps:	0

US-09-943-108a-2 (1-455) x AAS55331 (1-1368)

QY	1	MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys	20
DB	1	ATGCAATTTGAGGGTTATCAAGACGCTTCAAGCGACGATGCAAAAAATCGCTGTAAG	60
QY	21	GlyLysLeuThrGlnAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe	40
DB	61	GGTAACCTTACTAGCTGATATAAGATATATCGTGAAGTAAGATTAGCGTTACTTT	120
QY	41	GluAlaAspValAsnPhelysValValLysGluPheIleLysThrValSerGluArgAla	60
DB	121	CAGCGCTGACGTAACTTTAAGTGTAAAGAATTTATTAAACAGTATCAGACGCGCA	180
QY	61	LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln	80
DB	181	TTAGGTTCCGATGTAATGCAATCATTAACACAGCGCAACAAAGTTATTAAATAGTCAA	240
QY	81	AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro	100
DB	241	GATCAATTACCGACGTGATGGGTGGAGAAATACGTGATTAAATATGTCAAATAACCA	300
QY	101	ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys	120
DB	301	CCTACTCTGTTATGATGGTTGGTTTACAAGGCTCTGGTAAACCAACACATGCGAGTAA	360
QY	121	LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle	140
DB	361	TTAGCATTTATGATGGTAAATAACAAACAAACCTATGTTAGTTCCAGCAGATATT	420
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
DB	421	TATGTCACGACGATAAATCAATTAACACAGTAGGGAACAAATGATATCTCTGTA	480
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
DB	481	TACAGTGAAGGAGATCAAGTAAAGCAACAAATTTGTAACTAATGCAATTAACATGCT	540
QY	181	LysGluGlnHisLeuAspPheValIleLeuThrAlaGlyArgLeuHisIleAspGlu	200
DB	541	AAAGAAGAACATTTAGACTTTGTAATCATTCATACAGCAGGTGCGATTACATCGATGA	600
QY	201	AlaLeuMetAsnGluLeuLysGluValLysGlnIleAlaLysProAsnGluIleMetLeu	220
DB	601	CGATTGATGAACCAATTAAGAAGTAAAGACATTTGTAACCAACCAACCAATTTATGTTA	660
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln	240
DB	661	GTGTGCGATTCAATGACGGGTCAAGTGTCTCAATGTTGCGAATCTTTGACGATCAA	720
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla	260
DB	721	CTTGATGTCACAGTGTACCTTAACATAATAGATGGTGATACACGTGGTGGCAGCT	780
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu	280
DB	781	TTATCTATTGCTGCGTGACACAAAACCAATTAATTTGTTGGTATGAGTCAAAAGTTA	840
QY	281	AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
DB	841	GATGTTTAGAGCTATTCCTCTGACGTATGGCATCAGCTATTTTAGGTATGGGTGAT	900
QY	301	ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu	320
DB	901	GTGTTAAGTTTAATTGAAAAGCGCACCAAGATGTGATCAAGAAAAGCAAAAGATTTA	960
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln	340
DB	961	GAGAAAAAGATGCGTGAGTCATCGTTTACTTTAGATGATTTTATAGAACAACTTGATCAG	1020
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
DB	1021	GTGAAAAATCTAGGACCACTGGATGATTTATGAAAATGATTCAGGATGAATAAATG	1080

QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLeuAspHisIleLysAlaIlelle 380
 Db 1081 AAAGGGCTAGATAGCTTAATATGAGTGAAAGCAAAATTCATCAATTAATTAATGCGTATTC 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 CAGTCAATGACGGCGGTGAAAGAAACAATCCAGACACATGTAATGATATCAGTAAAG 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATGCTAAAGGGTCTGCTGCTCATTAACAAGATCAATCGTTTGATGAACAATTT 1260
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
 Db 1261 AACGATATGAAGAAATGATGAACAGTTCACCTGGTGGCGGTAAAGTAAAGGTA 1320
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 Db 1321 CGCAATCAATGCAAAATATGTAAAGGTATGAATTTACCGTTT 1365

RESULT 3
 AAS51783
 ID AAS51783 standard; DNA; 1365 BP.
 AC AAS51783;
 XX
 DT 13-FEB-2002 (first entry)
 DE Staphylococcus aureus DNA for cellular proliferation protein #200.
 XX
 KW Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Staphylococcus aureus.
 XX
 PN W0200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 PI WPI; 2001-611495/70.
 DR P-PSDB; AAU33924.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX Claim 27; Seq ID No 4365; 511pp; English.
 XX
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1365 BP; 518 A; 190 C; 285 G; 372 T; 0 other;
 SQ

Alignment Scores:

Pred. No.: 1,2e-193 Length: 1365
 Score: 2235.00 Matches: 447
 Percent Similarity: 98.90% Conservative: 3
 Best Local Similarity: 98.24% Mismatches: 5
 Query Match: 98.28% Indels: 0
 DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS51783 (1-1365)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTGAAAGGCTTATCAGAAGCGCTGCAAGCGACGATGCAAAAAATGCGTGAAG 60
 QY 21 GlyLysLeuThrGluAlaAspLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GGTAACTTACTGAAGCTGATTAAGATATGATGCGTGAAGATTAACGTTACTT 120
 QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
 Db 121 GAGGCTGACGTAAACITTAAGTGTGTAAGCAATTTATTAACACGATATCAGACGGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
 Db 181 TTAGGTTCCGATGTAATGCAATCATTAAACACGAGCAACAGTTATTAAATAGTTCAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GATGAATTAACGAAGTTGATGGGTGAGAAATATATCGATTATATGTCAAATAACCA 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCTACTGTTGTTATGATGGTTGGTTTACAGGTGCTGGTAAACAACAACACTCGAGTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTAGCATTATTGATGCGTAAAAAATACACAAAAACCTATGTTAGTTGCAGCAGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATCTCCAGCGCATTAATCAATTACAAACAGTAGGAAACAATAATGATTCCTGTA 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAATAATTGTAATGCATTAAACATGCT 540
 QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspIle 200
 Db 541 AAAGAAGAACAATTTAGACTTTGTAATCATATACAGCAGTCGATTACACATCATGATA 600
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 GCATTGATCAATGATTAAGAGATTAAGACATTCCTAAACCAACGAATTAATGTTA 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 661 GTTGTGATTCATGACGGGTCAAGATGCTGTCAATGTTGCAGAAATCTTTTGACGATCAA 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
 Db 721 CTTGATGTACAGGTGTTTACTTAACTAAATTAGATGTTGATACACGTGGTGGTGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleYspPheValGlyMetSerGluLysLeu 280
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 781 TTATCTATTGTTGGTGCACAAAACCAATTAAATTTGGTGGTATGAGTGAAGAGTTA 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 841 GATGGTTAGAGCTATCCATCGTAGAGTATGGCATCAGCTATTCAGGCGCTGGGTGAT 900
 QY 301 ValLeuSerIleLeuGlyLysAlaGlnAspValAspGlnGlyLysAlaLysAspLeu 320
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 901 GTGTCAAGGTTAATGAAAAGCGCACAGATGTGGATCAAGAAAAGCAAGAGATTTA 960
 QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 961 GAGAAAAGATGGCGAGTCATCTTTAGATGATTTTTAGAACACTTGTATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1021 GTGAAAAATTTAGACCACTGGATGATATTATTAATGATTCAGGTATGAATAAATG 1080
 QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLysHisIleLysAlaIleLe 380
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1081 AAAGGTCTAGATAAGCTTAATATGATGAAAAGCAAAATGATCATATTAAGCGATTATC 1140
 QY 381 GlnSerMetThrProIaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1141 CAGTCATGACGGCGGTGAAGAAACAAATCCAGACATGTAATGATCAGCTAAAGAG 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1201 CGTATTGCTAAAGGTTCTGGTGTTCATTACAGAAAGTCAATCGTTGATGAACAATTT 1260
 QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlySerLysGlyLys 440
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1261 AACGATATGAAGAAATGATGAACAACTCACTGGTGGCGGTAAAGGTAAAGAGTTAA 1320
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1321 CCACATCAATGCMAATATGTTAAAGGTATGATTTACCGTTT 1365

RESULT 4
 AAS54393
 ID AAS54393 standard; DNA; 1368 BP.
 XX
 AC AAS54393;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Staphylococcus aureus DNA for cellular proliferation protein #705.
 XX
 KW Antisense; ds: prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Staphylococcus aureus.
 XX
 PN WO2001170955-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HE;

XX WPI; 2001-611495/70.
 DR P-PSDB; AAU36534.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX Claim 27; Seq ID No 8030; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1368 BP; 520 A; 190 C; 285 G; 373 T; 0 other;

Alignment Scores:
 Pred. No.: 1-21e-193 Length: 1368
 Score: 2235.00 Matches: 447
 Percent Similarity: 98.90% Conservative: 3
 Best Local Similarity: 98.24% Mismatches: 5
 Query Match: 98.28% Indels: 0
 DB: 23 Gaps: 0

US-09-943-108A-2 (1-455) x AAS54393 (1-1368)

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1 ATGGCATTTGAAGCTTATCAGAACGCTGCAAGCAGCATGCAAAAATGGTGGTAAAG 60
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetArgGluValArgLeuAlaLeuPhe 40
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 61 GGTAAACTTACTGAAGCTGATATAAGATAATGATGCGTGAAGTAAAGATTACGTTACTT 120
 QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 121 GAGGCTGACGTAACTTTAAAGTGTAAAGAAATTTATTAACACATATCAGACGGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 181 TTAGGTTCCGATGTAATGCAATCATTAAACACAGGCAACAGTTATTTAAATAGTTCAA 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATATATCGATTATATGTCAAATAAGCA 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 301 CTAATCTGTTTATGATGGTGGTTTACAGGTGCTGGTAAACACACACACACAGTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 361 TTAGCATTTATGATGCGTGAAGAAATACAAACAAAAAACCCTATGTTAGTTCCAGCATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 421 TATCGTCCAGCAGCGATAAATCAATTACAAACAGTAGGGAACAAATGATATTCCTGTA 480

```
QY 161 TyrSerGluGlyAspGlnValLysProGlnInleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGAGATCAAGTAAGCCACAAATTTGTAATCATCAATAAATGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeaspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTGCTAATCATATACAGCAGGTCGATTACACATCGATCAA 600
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluLeuAlaLysProAsnGlnIleMetLeu 220
Db 601 GCATTGATGAATGATTTAAAGAGTAAGAAAGACATTTCTTAACCAACCAATTAATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATTCATGACGGTCAAGATCTGTCAATGTTGCAGAAATCTTTTGCAGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACCTTAATCAATAGATGATGATACAGTGTGTGTCAGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTAATCTATTCTGTCGACACAAACCAATTAATTTGTTGATGAGTGAATGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGCTTTAGAGCTATTTCATCTCTGAACGTATGTCATCACTGATTCAGGCTGCTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTCAAGGTTAAATTGAAAAGCGCAACAGATGTGGATCAAGAAAACCAAAAGATTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAGATGCGCGAGTCAATCTTACTTTAGATGATTTTAAACAACATTCATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGCAACCTGGATGATATTATTAATGATTCAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGTCTAGATAAGCTTAATATGATGAAGAAACAAATTCATATTAACCGGATTC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAsnProAsnThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACCGCGCTGTAAGAAACAAATCCAGACACATTCATGATGATCAAGTAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGGTTCTGTCCTTCATTACAAGAAGTCAATCGTTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysGlyLysGlyLys 440
Db 1261 AACGATATGAGAAATATGTAACCAATTCACCTGGTGGGTAAAGTAAAGGTAAAG 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACGGITT 1365
RESULT 5
ABQ70511/c
ID ABQ70511 standard; DNA; 2256 BP.
XX
AC ABQ70511;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #453.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
```

```
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX
PB (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
PR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
PS Claim 14; SEQ ID 3324; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB0711212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2256 BP; 533 A; 471 C; 420 G; 832 T; 0 other;

Alignment Scores:
Pred. No.: 1-45e-146 Length: 2256
Score: 1719.50 Matches: 331
Percent Similarity: 86.81% Conservative: 64
Best local Similarity: 72.75% Mismatches: 55
Query Match: 75.62% Indels: 5
DB: 24 Gaps: 1

US-09-943-108A-2 (1-455) x ABQ70511 (1-2256)
QY 1 MetaLAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1666 ATGGCATTTGAAGGACTAGCTGGAGACTCCCAAGAAACAATGACACAAATTCGCGCAAA 1607
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 1606 GGAAGAAGTAACAGAGCTGACGTAAGAAGAAATGATCGTGAGGTCTGCTCTACTT 1547
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 1546 GAAGCCGATGTTAACTTTAAAGTCGTTAAACAATTTATTAACAAGTAAAGCAAGCTGCT 1487
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 1486 GTGCGCGGAGCGTATGAAAGCCTTAACACCGCGCAACAAAGTTATCAAAATCGTCAA 1427
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 1426 GAAGAAGTAAACAGCTTAATGGCGGAGAGAAAGCAAAATTCGGAACACGCGACCGCCA 1367
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 1366 CCACCGGATTAATGATGTTAGTGTGTTACAGGGGCTGTGTAACACAGACTTCCGCGCAA 1307
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAspIle 140
```

```

Db 1306 CTTGCTAATTATTACCAAAATATAACCGTAAACCATTAAGTTCGTCACGACATATT 1247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 TyrArgProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspLeuProVal 160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1246 TACCGACCTCGCGCAATCAAAATATAGAAACACTTGGCAAGCAATAGATATGCGGTA 1187
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1186 TTTTCTTAGAGACCAAGTAAGCCGGTAGAATCGCCAAACAAAGCTATTGCTAAAGCA 1127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1126 AAAGAAGAACATTTAGATTATGTTATATCGATACAGGGGTGCTCTTCATATTGAAGAA 1067
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 AlaLeuMetAsnGluLeuLysGlnValLysGlnIleAlaLysProAsnGluIleMetLeu 220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1066 AGCCTAATGGATGAGTTAAACAAAGTGAAGAAATTCACACGCGCAACCGAAATTCGTGT 1007
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1006 GTTGTGATTCATGACTGGCGAAGATGCTGTACAGTGGCGCAAGCTTTCACGACAA 947
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 TTAGAAATACAGGTGCTGCTAAACCAATTAGACGGTGATACAGTGGTGGTCAGCG 887
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyLysValSerGluLysLeu 280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 886 CTTTGTGATTCGTCCTCGCTACGCAACCAATTAATTTATCGCGACTGCTGTAAGAAATG 827
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 826 GAAGCTCTAGAAATTTCCATCCAGATCGTATGCTCCAGAAATTCGCGATGGCGAT 767
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaIleAspLeu 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 766 GTGCTTCCCTGATTGAAAGCGCAACCGATGTAGATACAGAAAAATGAAGCTATG 707
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 706 GACAAAAAATGAAGACACACATGACGCTAGATGACTCTTAGAACAGTTACAAACA 647
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 GTAAACAAATGGGACCACTAGATGAATTAATTAATGATGCGCAGGCGCAACAAATG 587
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 586 AAAGGCTCGCAACATGAACGTCGATGATACAACTAGTGCATATTGAAGCAATATATC 527
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 AAGTCAATGACCAAAATGAAGACATCCAGATATCATCATCAATGCAAGTAGAGAAAA 467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 466 CGGATTGCACGTGTAGCGACGTCGGTTCAAGAAATTAACCGTCTCCCTAAACAAATTT 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 406 GGTGAATGAAGAAATGATGAAGCAAAATGACTGTGTGGCGCAAGGCAAAAGGCAAA 347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 441 ArgAsnGlnMetGlnMetLeuLysGlyMetAsnLeuProPhe 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 346 -----AATCCATTCCGCAATTTCAAAATGCAATTT 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
ABQ67194
ID ABQ67194 standard; DNA; 319630 BP.
XX AC
AC ABQ67194;
XX

```

```

DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #7.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
XX WPI; 2002-332479/37.
XX
XX New genomic sequences from Listeria species, useful for detection,
treatment and prevention of infection, also related polypeptides,
antibodies and modulators.
XX
PS Claim 5; SEQ ID 7; 180pp; French.
XX
XX The present invention relates to nucleic acid sequences
(CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
and primers for identification and/or detection of Listeria (e.g. as
contaminants in foods, or mutational analysis) and for analysis of
gene expression. Proteins encoded by the nucleic acid sequences can be
used to screen for compounds that modulate gene expression, replication
and pathogenicity of Listeria (potential therapeutic agents), also for
treating infections by Listeria, and are useful as immunogens in
anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format
directly from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 319630 BP; 105207 A; 55428 C; 66726 G; 92263 T; 6 other;
Alignment Scores:
Pred. No.: 6,7e-143 Length: 319630
Score: 1708.50 Matches: 329
Percent Similarity: 86.59% Conservative: 65
Best Local Similarity: 72.31% Mismatches: 56
Query Match: 75.13% Indels: 5
DB: 24 Gaps: 1
US-09-943-108a-2 (1-455) x ABQ67194 (1-319630)
QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 187924 ATGGCATTTGAAGGACTAGCTGGAAGACTCCAGAAAACATGAACAAATTCGCGGTAA 187983
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 187984 GGGAAAGTAAACGACGACTGACGTAAAGAAATGATGCGTGAAGTTCGCTTGCATTACTA 188043
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 188044 GAAGCGCATGTTAACTTAAAGTCGTTAAACAATTTATTAACAGTAGTAAGCAAGTGCA 188103
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysIleValGln 80
Db 188104 GTTGGCGGACGCTAATGAAAAAGCCTAACACCCGCTCAACAAGTAATTAATAATGTTCAA 188163
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
      :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::

```

Db	189164	GAAGAAATTACACAGCCTCATGGCGCGGAGAAAGAAATGAAAAATCGGACACAGCTGATGCCCG	1898222
QY	101	ProThrValValMetMetValGlyLeuGlnGlyValGlyLysThrThrThrAlaGlyLys	120
Db	188224	CCAAACCGTTATTATGATGGTAGTTTACAGAGAGCTGCTAAACACACACCTTCAGAGAAA	1898283
QY	121	LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle	140
Db	188284	CTFCGTAAATTATTACGTAAAAAATATPANTCGTAACCTTTTACTAGTCGCACAGATATT	1898343
QY	141	TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal	160
Db	189344	TACCGAGCTCGACCAATCAACAATAGAAACACTTGCAAAACATATAGATATGCCAGTA	1898403
QY	161	TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla	180
Db	189404	TTTTCTCTAGCGATCAAGTAGGCCAGTAGAAATCGCGAACACANGCTATCGCTAAAGCA	1898463
QY	181	LysGlnGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu	200
Db	189464	RAAGAGAACATTTAGATTATGTATTCATTCGATCAGCTGGCTGCTTTCATATCCAGCA	1898523
QY	201	AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu	220
Db	189524	ACTCTGATGGACGAATTAACAGTAGTGAAGAAATCGCTACGCCAACTGAAATTTTACTT	1898583
QY	221	ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln	240
Db	189584	GTAGTTGATTCATGACTGGGCAAGACGAGTAATGTGGCCAAAGCTTCAACGACAA	1898643
QY	241	LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla	260
Db	189644	TTAGAAATATCCGCGCTGTATTACAAAAATTAGACGGTGATACAGCTGGTGGGCGACA	1898703
QY	261	LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu	280
Db	189704	CTTTCCATCCGTTCACTCAGAGAAAACCAATCAAAATTCGTGCTACCGGTGAAAAAATG	1898763
QY	281	AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp	300
Db	189764	GAAGCAATCGAAACGTTCCATCCGATCGTAGTGGCTTCAAGAAATTCGCGCATGGGTGAT	1898823
QY	301	ValLeuSerLeuIleGluLysAlaGlnInAspValAspGlnGluLysAlaLysLeu	320
Db	189824	GTACTTCTCTTATTGAAAAAGCAACAACTGATGTAGATGCAGAAAAAATGAAAGCTATG	1898883
QY	321	GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnInLeuAspGln	340
Db	189884	GAACAAAAAAGAAAGACACAGCATGACCTTAGATGACTTCTTTGGACCAATTGCAACA	1898943
QY	341	ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet	360
Db	189944	GTAAAACAAAATGGGACCACTAGATGAACACTATAAAATGATGCCGGGCAACAAATG	1899003
QY	361	LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle	380
Db	189004	AAAGGCGCTGCACACATCAAGTAGATGATAAACAACCTCGCGTCACATCGAAGCGATAAT	1899063
QY	381	GlnSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys	400
Db	189064	AAATCCATGACCAAAAAACGAAAAAGATATPCGGGACATCATCAATCGCAGAGAGAAA	1899123
QY	401	ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe	420
Db	189124	CGAATTGCTGCTGGAGCGGCGGCCCAATTCAGAAATATCAATCGCTTCTCTAAACATTT	1899183
QY	421	AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLysLys	440
Db	189184	GCTGAATGAAAAAATGATGAACAAATGACTGGTGGGAGAAAAGSTAAAGAGGTAAA	1899243
QY	441	ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe	455
Db	189244	-----TAATCCATTCGCGCAATTTCAAAATGGGCAATT189277-----	

RESULT 7	
ABN92495	
ID	ABN92495 standard; DNA; 1014 BP.
XX	
XX	ABN92495;
XX	
DT	24-JUL-2002 (first entry)
XX	
XX	Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1958.
DE	
XX	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection
KW	antibacterial; gene therapy; gene; ds.
XX	
XX	Staphylococcus epidermidis.
OS	
XX	US6380370-B1.
PN	
XX	30-APR-2002.
PD	
XX	13-AUG-1998; 98US-0134001.
XX	
XX	14-AUG-1997; 97US-055779P.
PR	
XX	08-NOV-1997; 97US-064964P.
XX	
XX	(GENO-) GENOME THERAPEUTICS CORP.
PA	
XX	Doucette-Stamm LA, Bush D;
PI	
XX	WPI: 2002-381255/41.
DR	
DR	P-PSDB; ABP39950.
XX	
XX	Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT	polypeptide, useful for diagnosing and treating bacterial infections -
PT	
XX	Disclosure; SEQ ID 1958; 267pp; English.
PS	
XX	
XX	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC	given in ABP35124 to ABP3960. The S. epidermidis sequences have
CC	antibacterial activity and can be used in gene therapy. The sequences
CC	can also be used in the diagnosis and treatment of bacterial infections
CC	particularly S. epidermidis infections. The sequences can be used to
CC	screen for compounds able to interfere with the S. epidermidis life
CC	cycle or inhibit S. epidermidis infection.
CC	N.B. The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	USPTO web site.
XX	
XX	Sequence 1014 BP; 354 A; 156 C; 221 G; 283 T; 0 other;
SQ	
	Alignment Scores:
Pred. No.:	3,75e-130 Length: 1014
Score:	1534.00 Matches: 307
Percent Similarity:	97.29% Conservative: 16
Best Local Similarity:	92.47% Mismatches: 9
Query Match:	67.46% Indels: 0
DB:	24 Gaps: 0
US-09-943-108A-2 (1-455) x ABN92495 (1-1014)	
Qy	1 MetalApeGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
	: : : : : : : : :
Db	16 ANGCGATTGAAGGATTATCCGATCGCTTACAGCCAGCATGCAAAAATGCGTGATAA 75
	: : : : : : : :
Qy	21 GlyLysLeuThrGluAlaAspLeuLysIleMeKetArgGluValArgLeuAlaLeuPhe 40
	: : : : : : : :
Db	76 GGAAGAAGTACAGACGACATATTAACACGATGATCGCTGAAGTGAGATTAGCGTTATTG 135
	: : : : : : : :
Qy	41 GlnAlaAspValasnPhelysValValLysGluPheIleLysThrValSerGluArgAla 60
	: : : : : : : :
Db	136 GAAGCCGATGTTAACTTCAAGTTGTTTAAGGAATTGTTTAAGAATTGTTTCAGAACGAGCG 195
	: : : : : : : :

```

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValIleLysIleValGln 80
Db 196 CTAGGTTCTGATGTCAGTCACTTTACACCTGGCCACAGGTTATATAATCGTACAA 255
QY 81 AspGluLeuThrLysLeuMetGlyGlyValAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256 GAGACACTACTAGTTTATGTTGGTGAGAAATACCTCCATTAGATGGCAACAAACCA 315
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 316 CCAACTGTTGTCATGATGGTCGGCTTACAGGTGACAGGTAAACGACGACAGGTAAG 375
QY 121 LeuAlaLeuLeuMetAspLysLysValAsnLysProMetLeuValAlaAspIle 140
Db 376 TTGGCATATTATTCGCGTAAATAATATACAAATACCTTACTGTGGCAGAGATAT 435
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 436 TATCGTCCAGCTGCTATTGTATCAATTACAAACAGTAGTAAACAAATTGATATCCCTGTG 495
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 496 TATAGTCAAGGTGATCAAGTATCAACACACAAATTTGTAATGCTTTAAACATGCT 555
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 556 AAGAGACACATCTAGATTCCTTATCATGATGATACAGCTGCTGTTACACATGTGATGA 615
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 616 GCACCTTATGATGAGCTTCAAGAGTAAAGAAATCTCTAAACACGACGAAATATGCTT 675
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 676 GTTGTGTGATGATGACAGGTCAAGATGCTGTGAATGTGGCACATCATTTGATGACAA 735
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
Db 736 TTAGATGTTTCAGGTGTAACATTGACTAAATAGATGATACACGCGTGTGTCAGCA 795
QY 261 LeuSerIleArgSerValThrGluLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 796 CTTTCAATCCGTTCCGTTACCCAAAACCTATTAAATTTGTAGGTATGATGAGAAATG 855
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 856 GATGGTTTGAATATTATTCCTCGAAGCAATGCCCTCACGATTTTAGGTATGGGIGAT 915
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 916 GTTTTAAAGCTCATTTGAAAAGCGCAACAGATGTAGATCAAGAAAAGGCTAAAGATTTA 975
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAsp 332
Db 976 GAGAAGAAGATGAGAGATCATCTCTCTCTAAAT 1011

```

RESULT 8

AAS51308

ID AAS51308 standard; DNA; 1416 BP.

XX AAS51308;

AC AAS51308;

XX 13-FEB-2002 (first entry)

DE Enterococcus faecalis DNA for cellular proliferation protein #85.

XX Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

XX Enterococcus faecalis.

OS Enterococcus faecalis.

XX WO200170955-A2.

FN XX

```

PD 27-SEP-2001..
XX
EF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 25-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU33449.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
FS Claim 27; Seq ID No 3890; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1416 BP; 494 A; 228 C; 337 G; 357 T; 0 other:

```

Alignment Scores:

```

Pred. No.: 1.74e-119 Length: 1416
Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 23 Gaps: 2

```

US-09-943-108A-2 (1-455) x AAS51308 (1-1416)

```

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCTTTTGAGAGTTTACAAACCGCTACACAGGCAATGAGTAAATCCGTCGTAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAAAATTTCCCAAGCCGACGTAAGAAATGATCGAGAAATCCGTTGGCTTTATTA 120
QY 41 GluAlaAspValAsnPhelysValIlyLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAAGCCGACGTTAATTTTCAAGTGTGTCAAGATTTCCAAAAACGCGTCAGAGACGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTAGGATCGAAGTATTAGAAAGCTTATCACCAGCCCAAAATTTGTAATAATTTGTGAT 240

```


Score: 1418.50 Matches: 273
Percent Similarity: 77.94% Conservative: 86
Best Local Similarity: 58.96% Indels: 89
Query Match: 62.38% Gaps: 15
DB: 23

US-09-943-108a-2 (1-455) x RAS53127 (1-1434)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCTTTTTCAGAGTTTAAACAACCGCTACACAGGCAATCAGTAAATCCCGCTGAAG 60
QY 21 GlyLysLeuThrCluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAGAATTTCGAGAGCGAGCTAAGAATAATGATCGGAAATCCGTTGGCTTTATTA 120
QY 41 GluAlaAspValAsnPhelValValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAAGCCAGCGTTAATTACAAAGTGTCAAAAGTTTCACAAAACGCGTCAGAGAACGGCA 180
QY 61 LeuGlySerAspValMetClnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTAGAGTCGAAGATTTAGAAGCTTATCACCGCCCAACAAATGTGAATAATGTTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAAGAATTACGAACACGTTAGGTTTCAGAAACGTTGAACCTGAATAATCTCCAAAATC 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCGACAGTGATTATGATGACAGGTTTACAAGGGCGTGTAAACAACTTTTACTGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysIleLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTACGAAAACACTATGAAACCTGAAACCGCTCGCCCTTTTAATCGCTGGTGGAGTT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATTGATCAGTTGAAGGTTTAGGTCACAAATTAAGATTCCCGTT 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTGATATGGGACAGATCTAATCCAGTGGAAATTTGTCGTCAGGGTTAGCATAGCA 540
QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAAGAAATGATATGCTCTTAATGATACGCGCGCGCTTACACATTGACGA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCTTTAATGGACGAATTGAACAATAAAGAGTTGGCTAATCCCAATGAAATTCGTGA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATCGGATGAGGGGACAGATGCTCTACGTTGACAGATAGTTTAATGACAG 720
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGGAATTACTGGGTTCTTATTACCAAAATGACGCGGATCTCGTGGGGGGCGTGG 780
QY 261 LeuSerIleArgSerValThrClnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCAATTCGGGACATAACGGGCGCTCCGATTAAATTTGTCGGTCTGGTGAATAA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACCGATTAGAAATTTCCATCCGATCGTATGTCGAGTCGATCTCTAGGTATGGGGAC 900
QY 301 ValLeuSerLeuIleGlnLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGTTGACGCTAATTGAAACCGCAACAGATTACGATCAGAAAAGCAGAGACACTT 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspPheLeuGlnLeuAspGln 340

Db 961 GCTCAAAAATGAAGAAAACAGTTTTCAGCTTAACGATTTTCATGAGCAATGGATCAA 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTTATGGCATGGGACCGAATGAAGACTTATTAAATAATGATCCCTGATGAGTAACATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 1081 CTGGTATTGAAAATGTCAAAGTCGATCCAAAGATGTGGCAGCAACGGCGATGTC 1140
QY 381 GluSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CTATCAATGACCCCTGACAGACGTAAGATCTGATCTATTAAATCTAGTCGCGTGGC 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 AGAATTGCAGCTGGTTCAGGAATAAGTGTGTGAATCAATCTATGATTAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThr 431
Db 1261 AAGAATCCAAAATAATGTCACACAAATGTCCAAAGGGATATGAACATTCCTGGTATG 1320
QY 432 -----GlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMetGlnAsn 446
Db 1321 GATCAAAATCTAGTGGCGCGGTTAAAGCAAGTTAGTAAATG---GCCATGAATCGT 1377
QY 447 MetLeuLys 449
Db 1378 ATGATGAAG 1386

RESULT 10
AA12974
ID AA12974 standard; DNA; 6729 BP.
XX
AC AA12974;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:37.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN W09850555-R2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 406-409; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AA12938 to AA13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based

XX Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX Claim 4; SEQ ID NO 1847; 200pp; English.
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions.
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1136 BP; 338 A; 263 C; 326 G; 209 T; 0 other;

Alignment Scores:
Pred. No.: 4.18e-118 Length: 1136
Score: 1402.00 Matches: 270
Percent Similarity: 87.30% Conservative: 60
Best Local Similarity: 71.43% Mismatches: 48
Query Match: 61.65% Indels: 0
DB: 24 Gaps: 0

US-09-943-108a-2 (1-455) x ABK74556 (1-1136)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGGCATTGCAAGGATTACCGGCGGACTGCGAGCAACGATCTCAAAATCCGCGGAAA 60
QY 21 GlyLysLeuThrGluAlaAspLysLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGAAGAGTAAAGGCAAGACGCTCAAGAGATGATGCGCGAAGTCGCTTTAGCGCTCCTT 120
QY 41 GlnAlaAspValAsnPhelLysValLysGluPheLysThrValSerGluArgAla 60
DB 121 GAAGCGGACGTCATTTAAATGCTGAAGATTTGTGAAAAGGTAAAGTAAAGCAAGCT 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnValLysLysLeuValGln 80
DB 181 GTCGACAAAGAGCTTATGAAGAGCTGACGCGCGACAGCAGCTCATTAAGTGTAAAA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerLysMetSerAsnLysPro 100
DB 241 GAAGAGCTGACGAGCTGATGGCGGCGGAGGAAGCAAGATTGCGCTGCAAGCGTCOG 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
DB 301 CCGACCGTCATCATGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 121 LeuAlaLeuLeuMetArgLysLysLysLysLysLysLysLysLysLysLysLysLys 140

DB 361 CTTCGCAACCTGCTTCGCAAAAAACATAACCGCAATCCGCTGCTGCGGACGAGCATC 420
QY 141 TyrArgProAlaAlaAlaAsnGlnLeuGlnThrValGlyLysGlnLysLysLysLysLys 160
DB 421 TACCGCCGCGCGCGATCAAAACAGCTGGAACGCTCGGCAAGCAGCTGATATGCGGTA 480
QY 161 TyrSerGluGlyAspGluValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
DB 481 TTTTCATCTGGGTGACAGGTGATGAGATTCGCCAGCAGCGCCATCGAAAGGCG 540
QY 181 LysGluGlnHisLeuAspPheValLysLeuAlaGlyArgLeuHisLysLysLysLys 200
DB 541 AAAGAGATCATCAGACTGCTTCATCATGATACGCGGCGGACGCTTCATATGATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLysLysLysLysLysLysLys 220
DB 601 GAGCTGATGATGAGCTTGAAGAGTAAAGAACCGCTCAGCGGAGAGAGATTTCCCT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAsp 240
DB 661 GTCCTTGACCTCCATGACGGGTCAAGCGCTCAATGTTGCCAAAGCTTTAAGCAACAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspLysLysLysLysLysLysLys 260
DB 721 CTCGGTGTGACAGGTGCTCATCTTACGAAAGCTTGACGCGGATCTAGAGCGGAGCG 780
QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
DB 781 TTGTCAATCGCGCGCTGACAAACACGCGGATCAAAATTCGCGGAAATGGCGGAAAG 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgLysLysLysLysLys 300
DB 841 GATCGCTGGAGCGCTTCACTCTGACGCGATGCTTCGCGGATCTCGAATGGCGGAT 900
QY 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspValAspGlnGluLysLys 320
DB 901 GTGCTCACCTTATCGAGAAAGCCAGCCCAACAGTGAAGCAAGAAAGCGGAAAGCA 960
QY 321 GluLysLysMetArgGluSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
DB 961 GAGCAAAAAATGAAATATGAGCTTACGCTCGACGACTTTTGGAAACAGCTCGGACAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspLysLysLysLysLysLysLysLysLysLysLys 360
DB 1021 GTCCGCAATATGGCGCGCTTGAAGACCTGATTCAAATGATGCCGCGCGGCAAAATG 1080
QY 361 LysGlyLeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 378
DB 1081 AAAGGCTTAAAAAACGTGAAGGTTGAGTAAAAACAGCTGACGCTATATCGAAGCG 1134
RESULT 12
AAH90837 standard; DNA; 1605 BP.
XX AAH90837;
XX AC
XX DT
XX 02-OCT-2001 (first entry)
XX 2CFE 28 coding sequence.
XX Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
KW CFE; CEG; Conserved Essential Gene; bacterial infection;
KW antisense therapy; antibiotic resistance; ds.
XX Streptococcus pneumoniae.
XX OS
XX PN WO200149721-A2.
XX 12-JUL-2001.
XX 29-DEC-2000; 2000WO-US35604.
XX

PR 30-DEC-1999; 990S-0174089.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
 PI Tanassi JA;
 XX WPI; 2001-496721/54.
 XX
 XX Nucleic acids encoding conserved essential genes involved in bacterial
 PT replication which are potential targets for the treatment of antibiotic
 PT resistant bacterial infections -
 XX Claim 30; Fig 49; 380pp; English.
 XX
 CC The present invention relates to nucleic acids (AAH90701-AAH90918)
 CC encoding polypeptides (AAH01002-AAH0114), which are essential for the
 CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For
 CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic
 CC acids are useful for detecting the presence of proteins essential for the
 CC viability of a bacterial cell wall in samples such as cells, tissues,
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,
 CC and for detecting corresponding target nucleic acid molecules with
 CC complementary sequences. The nucleic acids are also useful for
 CC determining whether a genomic nucleotide sequence of interest is
 CC essential for viability of a bacterial cell or whether it resides within
 CC an operon, by integrating an exogenous nucleotide sequence comprising a
 CC portion of an open reading frame of the genomic sequence of interest
 CC (comprising 200-500 base pairs) into the genomic sequence of interest
 CC which confers a selectable phenotype to the cell, and determining cell
 CC viability with a selection agent such as chloramphenicol. The nucleic
 CC acids and proteins are also useful as vaccines and for treating bacterial
 CC infections with gene therapy and antisense therapy. The nucleic acids
 CC also enable identification of targets suitable for the treatment of
 CC antibiotic resistant bacterial infections.
 XX
 SQ Sequence 1605 BP; 484 A; 317 C; 393 G; 411 T; 0 other;

Alignment Scores:

Pred. No.: 9.54e-110 Length: 1605
 Score: 1312.00 Matches: 254
 Percent Similarity: 74.95% Conservative: 90
 Best Local Similarity: 55.34% Mismatches: 101
 Query Match: 57.70% Indels: 14
 DB: 22 Gaps: 2

US-09-943-108A-2 (1-455) x AAH90837 (1-1605)

QY 1 MetalloPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 DB 1 ATGGCATTTGAAGTTTAAACAGAACGTTTTCAGAACGCTCTTTAAATAATCTACGTAATAAAA 60
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
 DB 61 GGAATAATCTCTGAATCTGATGTCGAGGAGGACCAACCAACCAAGAGTATGCTGGCTGCTC 120
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
 DB 121 GAGGCGGACGTTGCTTGCCTGTTGTAAGGACCTTATCAAGAAAGTATGCTGAGCGTCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 DB 181 GTCGGGCATGAGGTCATGTGATCACTTAATCCTCGCAGACAGATATTATAAATCGTTGAT 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 DB 241 GAGGACTGACAGCGCTTTAGTTCTGTACGCGAGAAATATCAAGTCACCTAGATT 300
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 DB 301 CCAACCATCATCATGATGTTGTTTACAGGGGCTGTAAACACACCTTTGCTGTGTTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140

DB 361 TTGGCCACAACTCAGAAAGAAAGAAATCTCTTGTGATGCTTCGCGCGGATATT 420
 QY 141 TyrArgProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 DB 421 TATGTCGAGCTGCCATTTAGCAGCTTAAGACCTTGGGACACACAGATTGATGCTGCTC 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 DB 481 TTGCACTTGGACACAGAACTACACCTCTTGAGATTACGTCACAGGTTTGGACAGGCC 540
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
 DB 541 CAACTAATCATACAGCACTATCTCTGATTGATCTCGGGTCTGTTGCAGATTGATGAG 600
 QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGlnIleLeuLysProAsnGlnIleMetLeu 220
 DB 601 CTCCTCATGATGAGCTTCTGATGTGAACATTTGGCTCAACCAATGAATCTTCCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
 DB 661 GTCGTTGATGCTATGATTGTCAGGAACACCAATGTTCCGCGTGAGTTTAAATGCTCAG 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
 DB 721 TTGGAAGTGTACGGGTCTACCTTACCAAGATTGATGGCGATCTCGTGGTGGTCTGCT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
 DB 781 CTGCTGTGTCTACATCTCTGGAACCAATCAAGTTCTACTGGTACAGGTGAAGAATG 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 DB 841 ACGGACATTTGAACCTTCCACCACAGCCGATGCTAGCCGATCTCTTGGTATGGGGAT 900
 QY 301 ValLeuSerIleLeuGlnLysAlaGlnGlnAspValAspGlnLysLysLysAspLeu 320
 DB 901 ATGCTCACTTTGATTGAGAAAGCTTCTCAGGAATACGATGAACAAAGACCTTGAATG 960
 QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
 DB 961 GCTGAGAAGATGCGGAAACACCTTTGATTTTATGATTTCATCATCATATAGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 DB 1021 GTGCAAAATATGGCGCGATGGAAGACTTGTCTCAAGATGATTCAGATGATGCCAACAAT 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGlnLysGlnIleAspHisIleLysAlaIleIle 380
 DB 1081 CCAGCCCTTCAAAACATGAAGGTGGATGAACGCCAGATTGCTCGTAACGTCCTCATTTG 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 DB 1141 TCTTCGATGACACCTGNAGAGCGTGAAACCCACAGATTTGTTAAATCCAGCGCTGCGCGT 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
 DB 1201 CGTATTGCTGCTGTTCTGAAATACATTCGCGAAGTCAATAAATTCATCAAGACTTT 1260
 QY 421 AsnAspMetLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLys 440
 DB 1261 AACCGGCTAAACAGCTCATGCGAGGCTGTTATGCTGGG-----AsnLeuPro 454
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----GATATGATAAATATGATGAGCAATGGGATTAATCCAAATAACCTTCT 1350
 DB 1300 -----GATATGATAAATATGATGAGCAATGGGATTAATCCAAATAACCTTCT 1350
 RESULT 13
 ID AAS55519
 XX AAS55519 standard; DNA; 1572 BP.
 AC AAS55519;
 XX

Db 841 AGGACATTGAACCTCCACCCAGCCGATGCTAGCCGTATCCTTGGTATGGGGAT 900
 QY 301 ValLeuSerLeuLeuGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 ATGCTCACTTGATTGAGAAAGCTTCTCAGGAATACGATGAACAAAAAGCCCTTGAATG 960
 QY 321 GluLysMetArgGluSerPheThrLeuAspPheLeuGlnLeuAspGln 340
 Db 961 GCTGAGAGATGGCGAAGAACCTTGTGATTATGATTTCATGATCAATTAGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspMetLysMetLysMetLysMetLysMet 360
 Db 1021 GTGCAAAATATGGGCGGATGAGACTTGCTCAGATGATCCAGATGATCCCAACAT 1080
 QY 361 LysGlyLeuAspLysLysLeuAsnMetSerGluLysGlnLeuAspHisLeuAlaLeu 380
 Db 1081 CCAGCACTTCAAAACATGATGAGGTGATGAGCCAGATGCTGTAACGCTGCCATGTG 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLys 400
 Db 1141 TCTCGATGACCTGAGAGCTGAAACCCAGATTGTATATCCAAAGCGTCGCCGT 1200
 QY 401 ArgLeuAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATGCTGCTGTTCTGGAATACATTCGTCGAGCTCAATAAAATTATCAAGGACTTT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysGlyLys 440
 Db 1261 AACCAGGCTAAACAGCTCATGACGAGGTGTTATGTCTGG----- 1299
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
 Db 1300 -----GATATGATAAATGATGAGCAATGGGGATTATATCCAAATAACCTTCCT 1350
 RESULT 15
 ID AAH90724
 AC AAH90724 standard; DNA; 1572 BP.
 AC AAH90724;
 XX
 XX 02-OCT-2001 (first entry)
 DE CFE 28 coding sequence.
 KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
 KW CFE; CFE; Conserved Essential Gene; bacterial infection;
 KW antisense therapy; antibiotic resistance; ds.
 XX Streptococcus pneumoniae.
 OS WO200149721-A2.
 XX 12-JUL-2001.
 XX 29-DEC-2000; 2000WO-US35604.
 XX 30-DEC-1999; 99US-0174089.
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
 PI Thanassi JA;
 DR WPI; 2001-496721/54.
 DR P-PSDB; AA01025.
 XX Nucleic acids encoding conserved essential genes involved in bacterial
 PT replication which are potential targets for the treatment of antibiotic
 PT resistant bacterial infections -
 XX
 XX Claim 16; Page 200; 380pp; English.
 PS The present invention relates to nucleic acids (AAH90701-AAH90918)
 CC

CC encoding polypeptides (AA01002-AA01114), which are essential for the
 CC viability of a bacterial cell wall. The acronym CFE stands for "CFE For
 CC Expression", where CFE stands for "Conserved Essential Gene". The nucleic
 CC acids are useful for detecting the presence of proteins essential for the
 CC viability of a bacterial cell wall in samples such as cells, tissues,
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,
 CC and for detecting corresponding target nucleic acid molecules with
 CC complementary sequences. The nucleic acids are also useful for
 CC determining whether a genomic nucleotide sequence of interest is
 CC essential for viability of a bacterial cell or whether it resides within
 CC an operon, by integrating an exogenous nucleotide sequence comprising a
 CC portion of an open reading frame of the genomic sequence of interest
 CC (comprising 200-500 base pairs) into the genomic sequence of interest
 CC which confers a selectable phenotype to the cell, and determining cell
 CC viability with a selection agent such as chloramphenicol. The nucleic
 CC acids and proteins are also useful as vaccines and for treating bacterial
 CC infections with gene therapy and antisense therapy. The nucleic acids
 CC also enable identification of targets suitable for the treatment of
 CC antibiotic resistant bacterial infections.
 XX

SQ Sequence 1572 BP; 479 A; 298 C; 384 G; 411 T; 0 other;

Alignment Scores:
 Pred. No.: 3.26e-109 Length: 1572
 Score: 1306.00 Matches: 252
 Percent Similarity: 74.73% Conservative: 91
 Best Local Similarity: 54.90% Mismatches: 102
 Query Match: 57.43% Indels: 14
 DB: 22 Gaps: 2

US-09-943-108a-2 (1-455) x AAH90724 (1-1572)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTTGAAAGTTTAAACAGACGTTTACAGACGCTTTTAAATAATCTACGTAAAAA 60
 QY 21 GlyLysLeuThrGluAlaAspLysLysIleMetMetArgLysValArgLeuAlaLeuPhe 40
 Db 61 GGAAATATCTGATCTGATGTCAGAGGCAACCAAGAAATTCGCTTGGCCTTCTC 120
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheLeuLysThrValSerGluArgAla 60
 Db 121 GAGGCGGACGCTTGCCTTGCCTTGTAAAGGACTTTATCAAGAAAGTTCTGAGCGTGCA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGluValLysIleValGln 80
 Db 181 GTGCGGATGAGGTCATGATGATACACTTAATCTTCCGACACAGATTAATAATCTTGAT 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAGGAACGTACAGCCGCTTTAGGTCTGTACAGCGCAGAAATATCAAGTACCTAAGATT 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyThrThrThrAlaGlyLys 120
 Db 301 CCAACCATCATCATGATGTTGTTTACAGGGGCTGGTAAACACACCTTCTGGTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTGCCCAACAACCTCAAGAAAGAAATGCTGCTCTTGTGATGATTCGCGCGGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATGTCACGCTGCCATTGCCAGCTTAAGACCTTGGGACACAGATTGATGCTCCCTGTC 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TTTCGACTTGGACAGAGATACACGCTTGTGAGATTGTACGTCAGGCTTGGCAAGGCC 540
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
 Db 541 CAACATAATCATACGACTATGCTTGTGATGATGACTGCGGGCTGTTTCCAGATTGATGAG 600
 QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluLeuMetLeu 220

```
Db 601 CTCCTCATGATGAGTTCCTGATGAAAACATTCGCTCAACCAATGAAATCTTCCTT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValValAlaGluSerPheAspGln 240
Db 661 GTCGTGATGCTATGATGTTGAGGAGACGCCAATGTCGCGTGAGTTAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 TTGGAGTGACTGGGGTCATCCTACCAAGATTGATGGGATACTCGTGGTGGTCTGCT 780
QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGTCTGTCTGCATCTACTGGAACCAACCAATCAAGTTCACTGGTACAGGTCAAAAGATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACGGCATGAAACCTCCACCCAGACCGCATGCTAGCCGATATCCTGGTATGGGGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 ATGCTCACTTGATGAGAAAGCTTCTCAGGAATACGATGACAAAGAACCCCTTCAATG 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAGATCGCGAAACACACTTGATTTAATGATTCATCGATCAATTAGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGGCGATGGGAAGACTTGCTCAAGATGATTCAGGTATGCAAAACAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCCCTTCAACATGAGGTGGATGACGCCAGATTGCTCTGTAACGTCCCATTTGTG 1140
QY 381 GlnSerMetThrProAlaArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACACCTGAAGAGCGTGAAAACCCAGATTGTTAAATCCAAGCGCTGCGCGT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGTGCTGTGTTCTGGAAATACATTCGCGAAGTCAATAAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
Db 1261 AACCAAGGCTAACACCTCATCGAGGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGGATTATCCAAATAACCTTCT 1350
```

Search completed: February 25, 2003, 01:52:44
Job time : 421 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 01:43:57 ; Search time 81 Seconds

(without alignments)
1722.690 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFGLSLRLQATQAKMRGK.....GKGRNQMNLLKGNLFF 455

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09943108/runat_21022003_151631_1864/app_query.fasta_1.647
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09943108_@CGL_1_1_61 @runat_21022003_151631_1864 -NCPU=6 -ICPU=3
-NO_XLIPX -NO_MMALP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/POCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2274	100.0	1368	4	US-09-035-382-1
2	1534	67.5	1014	4	US-09-134-001C-1958
3	1306	57.4	7577	4	US-08-961-527-46
4	1305	57.4	792	4	US-09-035-382-3
5	1297	57.0	1569	2	US-08-923-772-1
6	1297	57.0	1569	4	US-09-385-287-1
7	1003.5	44.1	1098	4	US-09-221-017B-928
8	853	37.5	4804	4	US-09-066-047-6
9	592.5	26.1	2166	2	US-08-317-401B-3
10	573.5	25.2	2877	2	US-08-317-401B-1
11	432	19.0	4370	4	US-08-981-527A-20
12	430	18.9	1015	4	US-08-981-527A-7

13	426.5	18.8	1248	4	US-09-134-001C-1980	Sequence 1980, Ap
14	425	18.7	836	4	US-08-858-207A-180	Sequence 180, App
15	421	18.5	1251	3	US-09-007-476-1	Sequence 1, Appli
16	397.5	17.5	1278	2	US-08-986-963-1	Sequence 1, Appli
17	397.5	17.5	1278	2	US-08-986-963-3	Sequence 3, Appli
18	397.5	17.5	1293	3	US-09-007-484-1	Sequence 1, Appli
19	397.5	17.5	1293	4	US-09-309-682-1	Sequence 1, Appli
20	397.5	17.5	2764	2	US-08-986-963-4	Sequence 4, Appli
21	396.5	17.4	13121	4	US-08-961-527-126	Sequence 126, App
22	390	17.2	831	3	US-09-007-484-3	Sequence 3, Appli
23	390	17.2	831	4	US-09-309-682-3	Sequence 3, Appli
24	351	15.4	390	4	US-08-651-155B-4	Sequence 4, Appli
25	344.5	15.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
26	344.5	15.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
27	329	14.5	1046	4	US-09-221-017B-496	Sequence 496, App
28	287.5	11.8	29793	4	US-09-302-812-38	Sequence 38, Appli
29	267.5	11.8	29793	4	US-09-511-477-38	Sequence 38, Appli
30	267.5	11.8	29793	4	US-09-511-507-38	Sequence 38, Appli
31	258	11.3	802	4	US-08-998-416-483	Sequence 483, App
32	137.5	6.0	2085	2	US-08-668-128B-7	Sequence 7, Appli
33	137.5	6.0	2085	2	US-08-905-445-7	Sequence 7, Appli
34	136	6.0	1821	4	US-08-477-831C-1	Sequence 1, Appli
35	136	6.0	1885	4	US-08-477-831C-9	Sequence 9, Appli
36	136	6.0	1961	4	US-08-477-831C-8	Sequence 8, Appli
37	136	6.0	2968	4	US-08-477-831C-13	Sequence 13, Appli
38	136	6.0	3044	4	US-08-477-831C-12	Sequence 12, Appli
39	135.5	6.0	1896	4	US-08-477-831C-10	Sequence 10, Appli
40	132	5.8	3033	4	US-09-134-001C-2341	Sequence 2341, Ap
41	131	5.8	30549	4	US-09-134-001C-322	Sequence 322, App
42	130.5	5.7	5361	4	US-08-973-462-2	Sequence 2, Appli
43	130.5	5.7	6152	4	US-08-973-462-1	Sequence 1, Appli
44	130	5.7	11091	4	US-09-134-001C-2243	Sequence 2243, Ap
45	128	5.6	3837	4	US-09-724-517-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-035-382-1
; Sequence 1, Application US/09035382
; Patent No. 6284515
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES
; FILE OF INVENTION: AND POLYNUCLEOTIDES
; FILE REFERENCE: GM50035
; CURRENT APPLICATION NUMBER: US/09/035,382
; CURRENT FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: 60/057,890
; EARLIER FILING DATE: 1997-09-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-035-382-1

Alignment Scores:
Pred. No.: 2,03e-249 Length: 1368
Score: 2274.00 Matches: 455
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-943-108A-2 (1-455) x US-09-035-382-1 (1-1368)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGluAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTTGAAGCGTATCAAGACCGCTGCAGCGACGATGCAAAAATATCGTGGTAA 60
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

```
Db 61 GCTAACTTACTGAAGCTGATATAAGATAATGATCGCTGAAGTAAAGATTACGTTATTT 120
QY 41 GluAlaAspValAsnPhelYsValVallYsGluPhelLeYsThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAACATTTAAAGTGGTAAAGAAATTTATTAACACATGATCAAGACGGCCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleYsIleValGln 80
Db 181 TTAGGTTCCGATGTAATGCAATCATACACACGGGCAACAAGTTATTAATATAGTTCAA 240
QY 81 AspGluLeuThrYsLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAAATACATCGATTATATATGTCAAATAACCA 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTGTATGATGGTGGTTTACAAGGTGCTGGTAAACACAACACTCGAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATGCGTAAATAACACAAAACCTATGTTAGTTCACACAGATATT 420
QY 141 TyrArgProAlaLalLeuGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCATATAATCAATTAACAACAGTAGGGAACAAATTCATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuYsHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACARCAAAATTTGTAATAATGATTAACAAATGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACAATTTAGACTTTGTATCATTCATACACAGCGTCGATACACATCGATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGAACGAATTAAGAAGTAAAGAAATTTGTAACCAACCAACAAATATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTGTGCGATTCAATGACGGGTCAAGATGCTGTCAATGTCGAAATCTTTTACCATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGTGATGTCACAGGTGTACTTACTAATAATTTAGTGTGATACACGTCGTGGTCCACT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 TTATCATTTCTGTTGCGTGACACAAAACCAATTAATTTGTTGATGATGAGTGAAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetClyAsp 300
Db 841 GATGTTTAGAGCTATTCCATCTCGAAGCTATGCGATCAGCTATTTTAGGTATGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTTTAAGTTTAATGAAAAGCGCAACAAGATGTGGATCAAGAAAAACCAAAAGATTIA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAAGATGCGTGTAGTCATCGTTTACTTATAGTATGATTTTAGAACAACCTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGACCCACTGGATGATATATGAAAATGATTCACAGGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTTAGATAGCTTATATAGTGAAGCAAAATTCATATATTAAACGGATTATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
```

```
Db 1141 CAGTCAATGACCGCGCTGAAAGAAACAATCCAGACACATTGAATGTATCATCAAGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGATTGCTAAAGGCTGTGCTGTTCTATTACAAGAAGTCAATCGTTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLysLysLysLys 440
Db 1261 AACGATATGAAGAAATGATGAACAATCTACTGCTGGCGTAAAGTAAAGTAAAGTAAAGTAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACCGTTT 1365
RESULT 2
US-09-134-001C-1958
; Sequence 1958, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1958
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1958
Alignment Scores:
Pred. No.: 2,35e-165 Length: 1014
Score: 1534.00 Matches: 307
Percent Similarity: 97.29% Conservative: 16
Best Local Similarity: 92.47% Mismatches: 9
Query Match: 67.46% Indels: 0
DB: 4 Gaps: 0
US-09-943-108A-2 (1-455) x US-09-134-001C-1958 (1-1014)
QY 1 MetalAPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 16 ATGGCATTTGAAGGATTTATCCGATCGCTTACAAAGCCACGATGCAAAAAATGCGTGTAAA 75
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 76 GGAAGATGATACAGCAAGCAGATATTAACGATGATGCGTGAATGAGATTACGTTATTG 135
QY 41 GluAlaAspValAsnPhelYsValVallYsGluPhelLeYsThrValSerGluArgAla 60
Db 136 GAAGCCGATGTAACTTCAAAGTGTGTAAGGAATTTGTAAAGATGTTTCAGAACGACG 195
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 196 CTAGTTCCTGATGTCATGCAATCTTAACACCTGSCCAACAGGTTTAAAAATCGTAA 255
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 256 GAAGAACTTACTAGTTTAAAGGCTGGAGAAATATCTTCCATTAAGATGGCAACAAACCA 315
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 316 CCAACTGTTGTCATGATGCTCGCTTACAGGTGTCAGTAAACAGCACACACAGGTAA 375
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 376 TTGGCATTTAATGCGTAAATAATATACAAAAAACCTTTACTTGTGCGAGGATATT 435
```

```

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 436 TATGTCAGCTGCTATTGATTAACAACAGTAGTAACAACAAATGATATCCCTGTG 495
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 496 TATAGTGAAGTATCAGATACCAACACAAATGTTGAATAAGCTTTAAACATGCT 555
QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 556 AAAGAAGAATCATGATTCGTTATCATGATGATGATGATGATGATGATGATGATGAT 615
QY 201 AlaLeuMetAsnGlnLeuLysGlnValLysGlnIleAlaLysProAsnGlnIleMetLeu 220
Db 616 GCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 675
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 676 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 736 TTAGATGTTTCAGGTGTAACATGATGATGATGATGATGATGATGATGATGATGAT 795
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 796 CTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 856 GATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915
QY 301 ValLeuSerLeuIleGlyAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 916 GTTTAAGCTCTATTGAAGAAGCCCAACAGATGATGATGATGATGATGATGATGAT 975
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAsp 332
Db 976 GAGAAGAGATGAGAGAGTCACTTCTCTCTAAT 1011

```

RESULT 3

```

US-08-961-527-46/c
; Sequence 46, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7577 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-46
;
; Alignment Scores:
; Pred. No.: 4, 91e-138 Length: 7577
; Score: 1306.00 Matches: 252
; Percent Similarity: 74.73% Conservative: 91
; Best Local Similarity: 54.90% Mismatches: 102
; Query Match: 57.43% Indels: 14
; DB: 4 Gaps: 2
;
; US-09-943-108a-2 (1-455) x US-08-961-527-46 (1-7577)
;
; QY 1 MetaLapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
; Db 3033 ATGGCAATTCGTAATCTGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2974
;
; QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
; Db 2973 GGAAAAATCTGTAATCTGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2914
;
; QY 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
; Db 2913 GAGCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2854
;
; QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
; Db 2853 CTCGGCGATGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 2794
;
; QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
; Db 2793 GAGGATGACAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2734
;
; QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
; Db 2733 CCAACCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2674
;
; QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
; Db 2673 TTGCCCAACAACTCAAGAAAGAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2614
;
; QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
; Db 2613 TATGTCAGCTGCCATTGACCACTTAAGACCTTGGGACACACAGATGATGATGATGAT 2554
;
; QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
; Db 2553 TTTCACCTTGGAAACAGAGTACCAGCTCTTGCAGATTGTACGTTTGGAGCAAGCC 2494
;
; QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
; Db 2493 CAACTAATCATACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2434
;
; QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
; Db 2433 CTCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2374
;
; QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
; Db 2373 GTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2314
;
; QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
; Db 2313 TTGGAAGTGAAGTGGGTGATCCTTACCAGAGATGATGATGATGATGATGATGATGAT 2254
;
; QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

```


; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/923,772
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dickinson, Todd Q
 ; REGISTRATION NUMBER: 28,354
 ; REFERENCE/DOCKET NUMBER: GM10080
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-994-2252
 ; TELEFAX: 215-994-2222
 ; TELEX:

; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1569 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-923-772-1

Alignment Scores:

Pred. No.: 4,49e-138 Length: 1569
 Score: 1297.00 Matches: 251
 Percent Similarity: 74.51% Conservative: 91
 Best local Similarity: 54.68% Mismatches: 103
 Query Match: 57.04% Indels: 14
 DB: 2 Gaps: 2

US-09-943-108a-2 (1-455) x US-08-923-772-1 (1-1569)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGGCATTGAAAGTTTAAACAGACGTTTCAGAACGCTCTTTAAATAATCTACGTAAATAA 60
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GAAATAATCTGATCTGATCCAGAGCGCACCAACCAAGATTCGGTGGCTGCTC 120
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
 Db 121 GAGCCGACGCTGCCTTGCTGTGTAAGGACITTTATCAAGAAAGTTTCGGTGGCGTGA 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 Db 181 CTCGGCATGAGTCATTGATACACTTAATCTCGCAACAGATTATTAATAATCGTTGAT 240
 QY 81 AspGlnLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAGGACCTGACAGCCGTTTAGTCTGATACGCGCAGAAATATATCAAGTCACCTAAGATT 300
 QY 101 ProThrValValMetKetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCACATCATCATGATGTTGTTTACAAAGGGCTGGTAAACACACCTTTGCTGGTAA 360
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTGGCCCAACAACCTCAGAAAGAAATGCTGCTCTTGTGATGATTCCGGCGGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATCTCCAGCTGCCATTGACACGCTTAAGACCTTGGACACACAGATTGATGCTGCTC 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHsAla 180
 Db 481 TTTCACCTTGGACAGAAGTACCAGCTGTAGATTTAGCTCAAGGTTTGGAGCAAGCC 540
 QY 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGln 200
 Db 541 CAAACTAATCATACGACTATGCTTTGATTGATGATCTCGGGTCTGTTGCAGATTGATGAG 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 CTCTCATGATGACCTTCGGTGATGAAAGTATGGCTCAACCAATGAATCTGCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPheAspGln 240
 Db 661 GTCGTTGATGCTATGATTGGTGCAGAGCAGCAATGTTGGCGGTAAATGCTCAG 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 721 TTGGAAGTACTGGGGTCACTCTTACCAAGATGATGGTACTCTGTTGCTGCTGCT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
 Db 781 CTGCTGTTGCTCATCATCTGGGAACCAATCAAGTTCATGCTACAGTGAAGAAAT 840
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 ACAGATATCGAAACCTTCACCCAGACCGTATGCTAGCCGTATCTTGGCATGGGGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 ATGCTCACTTGTATGAGAAAGCTTCAGGAATACGATGACGAACAAAGCCCTTGAATG 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 Db 961 GCTGAGAGATGCGCGAACAACACTTTGATTTAATGATTTTCATCGATCAATTAGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTGCATAATATGGCGCGATGGAAGACTTCTCAAGATGATTCAGGTATGCCCAACAAT 1080
 QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
 Db 1081 CCACACTCTCAAAACATGAAGTGGTGAACGCCAGATTGCTGTAACGCGCCATTTGG 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 TCTTCGATGACATCTGAAGACGTAAGAACCCAGATTGTTAAATCCCAAGCGTGGCGGT 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGTATTGCTGCTGTTTGGAAATACATTCGTCGAAGTCAATAAATTCATCAAGGACTTT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLys 440
 Db 1261 AACAGCGCTAAACAGCTCATGCGAGGCTGTTATGCTCTGGG----- 1299
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
 Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGATTAATCCAAATACCTTCCT 1350

RESULT 6

US-09-385-287-1
 ; Sequence 1, Application US/09385287
 ; Patent No. 6350857
 ; GENERAL INFORMATION:
 ; APPLICANT: Black, Michael T.
 ; TITLE OF INVENTION: NOVEL fth
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dechert Price & Rhoads
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/385,287
 FILING DATE: 30-Aug-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/923,772
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Dickinson, Todd Q
 REGISTRATION NUMBER: 28,354
 REFERENCE/DOCKET NUMBER: GM10080
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-994-2252
 TELEFAX: 215-994-2222
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1569 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-385-287-1

Alignment Scores:

Pred. No.: 4,49e-138 Length: 1569
 Score: 1297.00 Matches: 251
 Percent Similarity: 74.51% Conservative: 91
 Best Local Similarity: 54.68% Mismatches: 103
 Query Match: 57.04% Indels: 14
 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-385-287-1 (1-1569)

QY 1 MetalApehGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 1 ATGCCATTGAAAGTTTACAGAACGTTTGCAGAACCTCTTAAACATCTACGTAACAAA 60
 QY 21 GlyLysLeuThrGluAlaAspLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 61 GAAAAATCTCTGAATCTGATCTCCAGAGCAACCAAGAAATTCGCTGGCCTCTC 120
 QY 41 GluAlaAspValAsnPhelLysValLysGluPheLeuLysThrValSerGluArgAla 60
 Db 121 GAGGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLeuValGln 80
 Db 181 GTCGGCATGAGGTCTATTGATACCTTAATCTCGCAACAGATTATTAATTCGTTGAT 240
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 241 GAGGAACCTGACAGCGCTTTAGTCTGTACGCGCAGAAATTTATCAAGTCACCTAAGATT 300
 QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 301 CCAACATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 QY 121 LeuAlaLeuLeuMetArgLysLysThrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 361 TTGGCCAAACAACTCAGAAAGAAAGAAATGCTGCTCTTGTGATGATGCGCGCGATATT 420
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 421 TATGCTCCAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
 Db 481 TTTCGACTTGAAACAGAGATACCAAGCTGTTGAGATTGCTACAGGTTTGAGCAAGCC 540
 QY 181 LysGluGluHisLeuAspPheValLysLeuThrAlaGlyArgLeuHisLysAspGlu 200
 Db 541 CAAACTATCAACAGACTATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
 Db 601 CTCCTCATGATGAGCTTCGTGATGTAAGATATTGCTCAACCAATGAATCTGCTT 660
 QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
 Db 661 GTCGTTGATGCTGATGATGCTGAGAACCAATGTCGCGTGAGTTAATGCTCAG 720
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
 Db 721 TTGGAAGTGTGCTGGGTTCCTTACCAAGATTGATGCTGATGATGCTGCTGGTGGT 780
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
 Db 781 CTGCTGCTGCTCACATCACTGGAAACCAATCAAGTTCATGCTACAGGTGAAAAATT 840
 QY 281 AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 841 ACAGATATCGAAACCTTCCACACAGACGATGCTAGCCGATGCTTGGCATGGGGAT 900
 QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 901 ATGCTCACTTGTGATGAGAAAGCTTCTCAGGAATACGATGACAAAGCCCTTGAATG 960
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
 Db 961 GCTGAGAAGATGCGGAAACACCTTTGATTTTAAATTCATGATTCATCAATTAGATCAG 1020
 QY 341 ValLysAsnLeuGlyProLeuAspLysPheMetLysMetIleProGlyMetAsnLysMet 360
 Db 1021 GTGCAAAATATGGGCGGATGGAAGACTTGTCAAGATGATCCAGTATGCGCCCAACAT 1080
 QY 361 LysGlyLeuAspLysLysMetSerGluLysGlnIleAspHisIleLysAlaIleLeu 380
 Db 1081 CCAGCACTTCAACATGAGGTGATGAGCCGAGATGCTCGTAAAGCTGCCATTGTG 1140
 QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
 Db 1141 TCTTCGATGACATCTGAAGAAGCTGAAACCCAGATTTGTTAAATCCAAAGCGTCGCGT 1200
 QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
 Db 1201 CGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLys 440
 Db 1261 AACAGGCTAAACAGCTCATGAGGCTGTTATGCTGGG----- 1299
 QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
 Db 1300 -----GATATGAATAAATGATGAAGCAATGGGATTAATCCAAATAACCTTCT 1350

RESULT 7

US-09-221-017B-928/c
 ; Sequence 928, Application US/09221017B
 ; Patent No. 6444799
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSS, Bruce C.
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1120
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER
 ; STREET: 755 PAGE MILL ROAD
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FASTSEQ for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066,047A
 ; FILING DATE: 24-Apr-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/044,869
 ; FILING DATE: 25-Apr-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Superko, Colleen
 ; REGISTRATION NUMBER: 39,850
 ; REFERENCE/DOCKET NUMBER: 106.941.156
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 526-6000
 ; TELEFAX: (617) 526-5000
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4804 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ; US-09-066-047-6

Alignment Scores:

Pred. No.: 9,27e-87 Length: 4804
 Score: 853.00 Matches: 174
 Percent Similarity: 63.57% Conservative: 100
 Best Local Similarity: 40.37% Mismatches: 147
 Query Match: 37.51% Indels: 10
 DB: 4 Gaps: 3

US-09-943-108A-2 (1-455) x US-09-066-047-6 (1-4804)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
 Db 3490 CTTAGTTTAACTCTTTAACCAAGGGGTTTCTTCGCGTCAAAAGGTTAAGTGGAAAG 3549
 QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
 Db 3550 CCGGAGATATCCACCAAGGATTTGATCTGTGTAATAGAGATATATCTCAGGCATTTGTTG 3609
 QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
 Db 3610 GATCGGATGTAAATCTGTGTGTGTGACCAATTTATAGAACCGTAAAAAGCAAGATC 3669
 QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
 Db 3670 GTAGGGCGGATGATGATTAAGGGGTCCTCCGGAGCAATGTGTCATTAAGCGTATAGAA 3729
 QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
 Db 3730 GAGTCTTTGATTGAAGTTTATGAGTAATGAGAGAGCGCTGTCATCTTAAGCGAAAGATT 3789
 QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
 Db 3790 COTCGAGTAATCATGATGTGTGGCTTCAAGGTGTGTGAGACTACTAACACACATAAAA 3849
 QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
 Db 3850 GTTGCACTGAGGTAA---AGAGAGATCTTAAACCCGTTGGTACGCTTTAGACGTA 3906
 QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
 Db 3907 TATCGTCTCGACGTCGAGACAGCTGAAGGTTTGGCGTGTGATGGATGTGTTATAGACAGT 3966
 QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180

Db 3967 CTTCCCATCGTTGAGGAGCAAAACCACTTCATATTCCGAAGCGTCTATGAGGGAAGCG 4026
 QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaClyArgLeuHisIleAspGlu 200
 Db 4027 AGGCTCAAAAGGCGACGATGTGTCTTTTGATACAGCGGGGCGCTTGCATATCAATCAG 4086
 QY 201 AlaLeuMetAsnGluLeuLysGluValIleAlaLysProAsnGluIleMetLeu 220
 Db 4087 GACATGATAGATGAGCTGAAGTGTGTAAGAGAGAGGATATCACCAGCTGAATTTGATTG 4146
 QY 221 ValValAspSerMetThrGlyCinAspAlaValAsnValAlaGluSerPheAspAspGln 240
 Db 4147 GTTGTAGACTCCTTAATGGGCAAGATGCCGTACTATGTCGCAAGTTCATCAATGAGGAG 4206
 QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
 Db 4207 TTAGGCATTACTGGGACGATCTTACCAGGCGGATGGTGATCCTTAGGGGTGGTCTATC 4266
 QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlyLysLeu 280
 Db 4267 TTGCTATGAGTTGGTTCCTGATGTCCTTAAGTTCATGTCACGGGAGAGAGCGCT 4326
 QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
 Db 4327 GAAGATTGGACGATTTCTATCCTGATAGAATGCTAGAGATGTTAAATATGGGAGAT 4386
 QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320
 Db 4387 GTCCCATCTCTTTAGAAAAGCGGTGAGACGGTTGGCAGGATACAAATTAATGAGCTA 4446
 QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
 Db 4447 CAGCGGAAGGCCAAGAGGTAATTCGATTGGATGATCTTGTATTTCAGCTGAAAGCT 4506
 QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMet----- 357
 Db 4507 TTGATATAAATGGTGTGTTATGCTTAATATAAGTTTATACCCGCTTTCGGTAACGAT 4566
 QY 358 -----AsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHis 375
 Db 4567 ATAAACCAAGTTCGGGGGATA-----CCTGATGACAGCAAGTGCATG 4614
 QY 376 IleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsn 395
 Db 4615 TACATTGGATTTAATCAATCAATGACGAAGCAGGAGGCGGATCCTGAGATACCTGAAT 4674
 QY 396 ValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArg 415
 Db 4675 GGTGCGAGGAAGCAAGGATAGCGAAGGGTGGGGAGTTAAGGTTGATGCTGTAATGCG 4734
 QY 416 LeuMetLysGlnPheAsnAspMetLysLysMet 426
 Db 4735 TTGCTAAGCAGTATATATCAGATGATTCGATA 4767

RESULT 9

US-08-317-401E-3
 ; Sequence 3, Application US/08317401E
 ; Patent No. 5922561
 ; GENERAL INFORMATION:
 ; APPLICANT: Thompson, Sheryl Ann
 ; APPLICANT: Yaver, Debbie Sue
 ; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
 ; TITLE OF INVENTION: ASPERGILLUS NIGER
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
 ; STREET: 405 Lexington Avenue, Suite 6400
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM: disk
 ; MEDIUM TYPE: Floppy

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,401E
; FILING DATE: 03-October-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4248.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2166 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; US-08-317-401E-3

```

```

Alignment Scores:
Pred. No.: 1,24e-57 Length: 2166
Score: 592.50 Matches: 156
Percent Similarity: 50.11% Conservative: 75
Best Local Similarity: 33.84% Mismatches: 211
Query Match: 26.06% Indels: 19
DB: Gaps: 9

```

US-09-943-108a-2 (1-455) x US-08-317-401E-3 (1-2166)

```

QY 1 MetalalpheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 86 ATGGCTTCAGGATCGCGGGCGAATCAACGCGCGCAATGACGTACTCGCTCC 145
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 146 AACAATTTGGACGAGAGGCGCTTTGATGACATGATTAAAGAGATCTCGCGCGCTCGTG 205
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 206 TCGCGGAGCTACAGCTCCGCGCTGTCAGTCCCTCCGCAAGTCCATCAACCGCGCTC 265
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 266 AACTTTGCTCTCTCTCTCGCGCGGTGAACAAGAGCGTTTGTATCAAAAGCGGCTCTC 325
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 326 GATGAGCTGTTTCCCTGGTGTATCCCATGCGGAGCGCTTCCGCGCCCAAGAGGCGCG 385
QY 101 ProThrValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 386 TCAACGCTGATCAGTCTCGTGGCTGCGAGGCTGCGGTAAACACCACTGTACCAAG 445
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLys-----LysProMetLeuValAla 137
Db 446 CTGGCC-----CGGCATCCAGATGCGCGGCTTCAAGACTGCGCTCGTGT 493
QY 138 AlaAspIleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAsp 157
Db 494 GCGGATACCTCCGCTGGTGTGTTTCGACAGCTGAAGAGAGATGCCAAGAGCCAG 553
QY 158 IleProValTyrSerGluGlyAspGlnValLysProGlnIleValThrAsnAlaLeu 177
Db 554 ATCCCTACTAGGTAGCTTGACCAACCGACCGCCCGCCATGTAGCAGCGCGAGGTGTG 613
QY 178 LysHisAlaLysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHis 197
Db 614 GCCAAGTTCAAGAGAGCGTTTCGAAATCATCATCGTCGATACCATCGTGTGTCACAAAG 673

```

```

QY 198 IleAspGluAlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlu 217
Db 674 CAGGAAGAAGAGCTTTTCACCGAAATGACCGACATTCACCGCGCGCTCACCOCGACGAG 733
QY 218 IleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPhe 237
Db 734 ACCATCTCTGCTTCGACGACGACCATCGGTGCTGAGGTCCCGAAGCCCGCTCTGCTTC 793
QY 238 AspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGly 257
Db 794 AAGGCCACCGCAGACTCGGAGCATCATCATCACAAGACGATGGTCACGCCGAGGT 853
QY 258 GlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSer 277
Db 854 GGTGTGCTATTTCGCGCGTCGCGCACACACTCCCATTTACTACCTCGGTACCGGT 913
QY 278 GluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGly 297
Db 914 GAGCACCTGATGGATCTGGAGCGCTTTGAGCGGAGGCGCTTCATCCAGAAGCTCTCGGT 973
QY 298 MetGlyAspValLeuSerLeuGluLysAlaGlnGln---AspValAspGlnGluLys 316
Db 974 ATGGCGATATGGCTGCTGCTGATAGCAGCGTACAAAGCGGTGACCAAGGACTCTGCTCC 1033
QY 317 AlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlu 336
Db 1034 GCCAAGGAACCTCAAGACATCTCCGAGGTATCTACAGCTGCTGCTGCTCCGCGAG 1093
QY 337 GlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGly 356
Db 1094 AACATCACTCCATCATGAGATGGAGCGCTCTCCAAAGCTCTCCGCGATGATCCCGGT 1153
QY 357 Met----AsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHis 375
Db 1154 CTATCCAACTGACCGCGGTCTCGATGACGAGACGCGCTCCATGAAG---CTCCGTCGC 1210
QY 376 IleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsn 395
Db 1211 ATGATCTACATTCGACATGATGACGCGCGCGGCTCCATGATGCTGATGCTGATGCT 1267
QY 396 Val-----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
Db 1268 TTGCTGCAACAACCGACGCGCATGCTCGTATCGCTGCGGAGCGGTACCACTGCGC 1327
QY 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGlnPheThr 431
Db 1328 GAAGTGGAGACCTGCTCTCCAGCACCGCATGATGGCGCGCATGCGCAAGCGGTGTC--- 1384
QY 432 GlyGlyGlyLysGlyLysArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db 1385 ---GTTGACAGAGAGAGAGATGACAGCGC-----GCCAGACATGCTCAGGCGGT 1435
QY 452 Asn 452
Db 1436 AAC 1438

```

RESULT 10

```

US-08-317-401E-1
; Sequence 1, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Yaver, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```


QY 109 LeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuMetArgLysLys 128
 Db 400 GTAACGCGTGGGAAACACACAGATCGGAGAGCTTCTCTATAAATG--AAACA 456
 QY 129 TyrAsnLysLysProMetLeuValAlaAlaAspLeuYrgrProAlaAlaLeuGln 148
 Db 457 GAAGAAATCTGTGTACTTGGCGGAGACACTTTTATAGCGGAGCCATTGAACAG 516
 QY 149 LeuGlnThrValGlyGlnLeuAspLeuProValTyrSerGluGlyAspGlnValLys 168
 Db 517 CTGGAGATGCGGAGAGCGTACAGAGTGCCTGTCATTAAAGCAGCGCAGAGCCGAT 576
 QY 169 ProGlnGlnLeuValThrAsnAlaLeuLysHisAlaLysLeuLutisLeuAspPheVal 188
 Db 577 CCGGGCGTGTCACTACGATGCTGTTCATGCTCGAAAGCAAGAAATCGCATATTA 636
 QY 189 IleLeuAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLysGlu 208
 Db 637 ATTGTGATACGGCAGCGCGCTCCAAACAAAGTAAATCTCATGAAGAGCTTGAAAAA 696
 QY 209 ValLysGluIleAlaLys-----ProAsnGluIleMetLeuValVal 222
 Db 697 GTAACCGTGTATCGAAGAGAGTCTCTGAGCTCGCATGCGATGAGTGTGCTTGCCTT 756
 QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
 Db 757 GATGCCAGCGCGCCAAATGATGGCTCAGCAAAAGAAATTTCTTAAGCAACAAT 816
 QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
 Db 817 GTTACCGCATTTTAAACGAGCTTGACGGTACGGCAAAAGCGGTATCGTCTTGG 876
 QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
 Db 877 ATTCCGACGAGCTTCATCCCGGTAAACTAGTACGCTTTAGCAGAGAAAGTTGATGAC 936
 QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
 Db 937 CTCAGGAATTGTATCCAGAA-----TCCTATGTGTACGGACTC-----TTTCA 981
 QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLys 316
 Db 982 GATTAGTGGAAAGGCCGAC-----GATTAGAGAAAA 1014

RESULT 13
 US-09-134-001C-1980
 ; Sequence 1980, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 1980
 ; LENGTH: 1248
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-1980

Alignment Scores:
 Pred. No.: 4,2e-39 Length: 1248
 Score: 426.50 Matches: 103
 Percent Similarity: 54.57% Conservative: 76
 Best Local Similarity: 31.40% Mismatches: 122
 Query Match: 18.76% Indels: 27
 DB: 4 Gaps: 8

US-09-943-108A-2 (1-455) x US-09-134-001C-1980 (1-1248)
 QY 8 GluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLysLeuThrGluAlaAsp 27
 Db 298 CAAACTTCCCAAGCAAGTAAATAATTAATTCGTGATATAGAAAAGTTGACGAAGAT 357
 QY 28 IleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAlaAspValAsnPheLys 47
 Db 358 TTCCTCGAAGCTCTGGAGAAATG-----CTATTACTGGGACGTTGGTTTAAAT 408
 QY 48 ValValLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67
 Db 409 ACCGTATTGAAATTAACATGATGAGCTACGTACAGAAAGCACAAAGAGCTAATATACAAGAA 468
 QY 68 SerLeuThrProGlyGlnGlnValIle--LysIleVal-----GlnAsp 81
 Db 469 ACAGAGACTTAAGAGAGGTTTAGTTGAGAAGATTGTAGAAATCTATCATCAAGAGGAC 528
 QY 82 GluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProPro 101
 Db 529 GATTATTCTGAAGCAATGAATATTGAAGATGGACGTTTAAAT----- 570
 QY 102 ThrValValMetMetValGlyLeuGlnGlyValGlyLysThrThrAlaGlyLysLeu 121
 Db 571 --GTCATACTGATGGTGTGTAATGCTGCGCAAAACACAAATTTGGTAAATTA 627
 QY 122 AlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyr 141
 Db 628 GCTTATCTTATCAACAGAA--GGTAAAAAGTAAAGTTAGCCGCTGATACCTTC 684
 QY 142 ArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr 161
 Db 685 AGAGCTGGAGCAATTCACAAATTAACCTCTGGGAGAACGTTGTGTGTAAGTTGTG 744
 QY 162 SerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLys 181
 Db 745 AGTCAAAACGAGAGTCTGACCTCGCAGCAGTAGTATGATGCGGATTAATGCGCAAA 804
 QY 182 GluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGluAla 201
 Db 805 AATAAGCGGTAGATATTTAATTTGTGATCTGCAGGACGCTTGCAAAATAATCTAAT 864
 QY 202 LeuMetAsnGluLeuLysGluValLysGluIleAlaLys-----Pro 215
 Db 865 TTAATGCAAGAGTTAGATAAAATGAAACGTTGATTAATCTGTCGCAATACCTGATGCCGCC 924
 QY 216 AsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlu 235
 Db 925 CATGAAGCTTTATTATGCTTGGATGCAACACTGCTCAAAATGCACTTTCACAGCGGT 984
 QY 236 SerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThr 255
 Db 985 TCATTTAAGGAAGTTACAATGTCTCAGGTATAGTTTAACTAATATAGACGCTACTGCT 1044
 QY 256 ArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGly 275
 Db 1045 AAAGGGGTATTGTATTAGCAATTCGAAATGAGTTACACATTCACAGTTAAATATGTTGT 1104
 QY 276 MetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIle 295
 Db 1105 TTAGCGCAAAAGTGGATGACTTACACCGCTTAACTCTCTGAA-----AGCTATGTA 1155
 QY 296 LeuGlyMet--GlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314
 Db 1156 TATGGATTATTTGCTGATATGATAGACAAATGAGATATTCTCGAAGAAATCTCTAGA 1215
 QY 315 GluLysAlaLysAspLeuGluLys 322
 Db 1216 AATTCATCGTTGAATCTGAAGAA 1239
 RESULT 14
 US-08-858-207A-180


```
Db 349 TTTTGTGAAGCTTTAGAGAAATG-----TTATCACTGCAGACGTCGGTTTTTAAT 399
Qy 48 valValLysGluPheIleLysThrValSerGluArgAlaLeuGlySerAspValMetGln 67
Db 400 ACAGTG-----ATCAGCTTAACGTGAAGAATTAAGTATGGAAGCACAACACGCGT 447
Qy 68 SerLeuThrProGlyGlnValIleLysIleValGlnAspGluLeuThrLysLeuMet 87
Db 448 AATATTCAGATACATGAGATTTCGCGTGAAGTCATGTTGAAAAAATCCTAGAGATTAC 507
Qy 88 GlyGlyGlu-----AsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
Db 508 CATCAAGAAGATGATTAATTCAGAAGCTATGAACATAGAAGATGCTGTTAAATGTCATT 567
Qy 105 MetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
Db 568 TTAATGGTGGGTGTAATGTTGTTGTAACACACAAATTCGAAAAATTAGCT---TAC 624
Qy 125 MetArgLysLysTyrAsnLysLysPrometLeuValAlaAlaAspIleTyrArgProAla 144
Db 625 CGATATATAATCGAAGGTAAAAAGTAATGTTAGCTGGGGCGGATACTTTAGACGGGGT 684
Qy 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGly 164
Db 685 GCTATTGATCAATTCAAAAGTTGGGGCGAAGCTGTTGCTGTAGATGTAATTAAGCCAAAAGT 744
Qy 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
Db 745 GAAAGTTCGATCCAGCTGCTGTTATGTATGATGCGATTATCCCGCTAAAAAACAAAGGT 804
Qy 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
Db 805 GTTGATATTATTAATCTGTGTATCCGCTGGACGCTTACAAAAATAAACAAATCTAATGCAA 864
Qy 205 GluLeuLysGluValLysGluIleAlaLys-----ProAsnGluIle 218
Db 865 GAATTAGAAAAAGTTAAGCGTGTAATTAATCGAGCAGTCCAGATGGCGCTCATGAAGCA 924
Qy 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 925 TTACTATGTTAGATGCTACAACTGCTCAGATGCTGATGCTGCTCACAAGCTAGAACTTTAA 984
Qy 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGly 258
Db 985 GAAAGTAACAAATGTTACAGGTATTGTTATTAACGAAATTAGATGTCACAGCCAAAGTGCT 1044
Qy 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlu 278
Db 1045 ATCGTATTAGCCATTCGTAATGAATTGCACATTCACAGTTAAATATATGAGTTTAGTGAG 1104
Qy 279 LysLeuAspGlyLeuGluLeuPheHisProGluArg-----MetAlaSerArg 294
Db 1105 CAATTAGATGACTTACAACTTAACTCCCTGAAGCTTATGCTACGGCTTATTCGCTGAT 1164
Qy 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314
Db 1165 ATGATTGAACAAATATGACGAATAACAAACAGTTGAAAT---GATCAAAATTGTAACAGAA 1221
Qy 315 GluLys 316
Db 1222 GAAAG 1227
```

Search completed: February 25, 2003, 03:14:46
Job time : 109 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:59:07 ; Search time 89 Seconds
(without alignments)
2871.303 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274
Sequence: 1 MAFELSERLOATQMKRGK.....GKGGKRNQNMMLKGNLPPF 455

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
-Q=/cgn2.1/USFTO_spool/US09943108/runat_21022003_151633_1953/app_query.fasta_1.647
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USFR=US09943108 -CGN 1 1 80 -runat_21022003_151633_1953
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_XMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:*
1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2261	99.4	1368	10	US-09-815-242-8968 Sequence 8968, Ap
2	2235	98.3	1365	10	US-09-815-242-4365 Sequence 4365, Ap
3	2235	98.3	1368	10	US-09-815-242-8030 Sequence 8030, Ap
4	1418.5	62.4	1416	10	US-09-815-242-3890 Sequence 3890, Ap

5	1418.5	62.4	1434	10	US-09-815-242-6764 Sequence 6764, Ap
6	1417.5	62.3	6729	10	US-09-070-927A-37 Sequence 37, Appl
7	1402	61.7	1136	10	US-09-974-300-1847 Sequence 1847, Ap
8	1307	57.5	1572	10	US-09-815-242-9156 Sequence 9156, Ap
9	1306	57.4	1395	10	US-09-815-242-9427 Sequence 9427, Ap
10	1297	57.0	1569	10	US-09-814-041A-1 Sequence 1, Appli
11	1168	51.4	1641	9	US-09-738-626-2255 Sequence 2255, Ap
12	1164.5	51.2	1362	10	US-09-815-242-6166 Sequence 6166, Ap
13	1158	50.9	721	10	US-09-815-242-1949 Sequence 1949, Ap
14	1150	50.6	1374	10	US-09-815-242-7850 Sequence 7850, Ap
15	1138	50.0	1389	10	US-09-815-242-6868 Sequence 6868, Ap
16	1133.5	49.8	1506	10	US-09-815-242-9991 Sequence 9991, Ap
17	1006	44.2	640681	10	US-09-790-988-1 Sequence 1, Appli
18	854	37.6	1347	10	US-09-815-242-7465 Sequence 7465, Ap
19	847	37.2	521	10	US-09-815-242-1985 Sequence 1985, Ap
20	847	37.2	521	10	US-09-815-242-2013 Sequence 2013, Ap
21	764	33.6	468	10	US-09-815-242-3070 Sequence 3070, Ap
22	560	24.6	1440	9	US-09-938-842A-877 Sequence 877, App
23	558	24.5	330	10	US-09-815-242-2247 Sequence 2247, Ap
24	432	19.0	1494	10	US-09-815-242-6281 Sequence 6281, Ap
25	426	18.7	292	10	US-09-815-242-3369 Sequence 3369, Ap
26	418	18.4	1251	10	US-09-815-242-8028 Sequence 8028, Ap
27	416	18.3	948	10	US-09-815-242-4244 Sequence 4244, Ap
28	410	18.0	987	10	US-09-974-300-1852 Sequence 1852, Ap
29	406	17.9	15363	10	US-09-070-927A-279 Sequence 279, App
30	404.5	17.8	1329	10	US-09-815-242-6807 Sequence 6807, Ap
31	404	17.8	1368	10	US-09-815-242-7684 Sequence 7684, Ap
32	403.5	17.7	939	10	US-09-815-242-3937 Sequence 3937, Ap
33	397.5	17.5	1278	10	US-09-815-242-9165 Sequence 9165, Ap
34	397.5	17.5	1293	10	US-09-827-663-1 Sequence 1, Appli
35	396.5	17.4	1290	10	US-09-815-242-9491 Sequence 9491, Ap
36	395.5	17.4	1245	10	US-09-815-242-6997 Sequence 6997, Ap
37	390	17.2	831	10	US-09-827-663-3 Sequence 3, Appli
38	376	16.5	1530	9	US-09-738-626-2259 Sequence 2259, Ap
39	276	12.1	643	10	US-09-815-242-2506 Sequence 2506, Ap
40	268.5	11.8	610	9	US-09-736-457-1094 Sequence 1094, Ap
41	268.5	11.8	610	9	US-09-902-941-1094 Sequence 1094, Ap
42	268.5	11.8	610	9	US-09-849-626-1094 Sequence 1094, Ap
43	267.5	11.8	29793	10	US-09-973-451-38 Sequence 38, Appl
44	193	8.5	405	10	US-09-960-352-13494 Sequence 13494, A
45	189	8.3	420	10	US-09-960-352-14727 Sequence 14727, A

ALIGNMENTS

RESULT 1
US-09-815-242-8968
; Sequence 8968, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

;
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8968
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8968

Alignment Scores:

Pred. No.: 5,28e-245 Length: 1368
Score: 2261.00 Matches: 452
Percent Similarity: 99.78% Conservative: 2
Best Local Similarity: 99.34% Mismatches: 1
Query Match: 99.43% Indels: 0
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-815-242-8968 (1-1368)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTGAGGGTTATCAGACCCCTTGCAGCGCAGCATGCCAAAATGGTGGTAA 60
Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAAACTTACTGAAGCTGATATAAAGATAATCATGCGTGAAGTAAGATTAGCGTTACTT 120
Qy 41 GluAlaAspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGCGTCGACGTAACCTTTAAAGTGGTAAAGAAATTTATTAACACAGTATCAGACCGCA 180
Qy 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTCCGATGTAATGCATCATTAACACCGGCGCAACAGTTATTAATAGTCA 240
Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGATTAACCCAGTTCATGGGTGGAGAAAATAGTCGATTAATATGTCAAAATAAAC 300
Qy 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTTATGATGGTTGGTTTACAGGGTCTGGTAAACACACACTGCAGTAA 360
Qy 121 LeuAlaLeuMetArgLysLysTyrAsnLysLysPrometLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATTCGCTTAAATAATACAAAAAACCTATGTTAGTTCAGCAGATATT 420
Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATAATCAATTACAAACAGTAGGGAACAAATATGATATCTCTGA 480
Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGGAGATCAAGTAAAGCCACACAAATTTGTAATATGCTTAAACATGCT 540
Qy 181 LysGluGlnHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTTGAATCAATTGATACAGCAGGTGCGATTACACATCGATGAA 600
Qy 201 AlaIleMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTGATGAACGAATTAAGAGAGTAAAGACATTGCTTAAACCAACCAAGAAATATGTTA 660
Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGATTCATCACTGACGGGTCAAGATGCTGTCAATGTTGACAGATCTTTTTCAGATCA 720

Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTGTGATGTCACAGGTGTACCTTAACCTAAATTAGTGGTATACACAGTGGTGGTCACT 780
Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
Db 781 TTATCTATTCTGTCGTGACACAAAAACCAATTAATTTCTGTGTATGAGTGAAGTTA 840
Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 GATGGTTTAGAGCTATTCCATCTGAAGTATGGCATCAGCTATTTTAGGTATGGGTGAT 900
Qy 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGTTAAGTTTAATTTGAAAAGCGCAACAGATGTGGATCAAGAAAAACAAAGATTTA 960
Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnLeuAspGln 340
Db 961 GAGAAAAAGATCGGTGAGTCACTTACTTTAGTATGATTTTATAGACACACTTGTATCAG 1020
Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATCTAGGACCATCGGATGATATTAATAAATGATCCAGTATGATATAAATG 1080
Qy 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGGCTAGATAAGCTTAATATGATGTAAGCAAAATTCATCATATTAAACGATATC 1140
Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGAGCGCGGTGAAGAAAAATCCAGACACATTCAGTATGATCACTGATAAAG 1200
Qy 401 ArgIleAlaLysLysSerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATGCTAAAGGTCTGCTCTTCATTTACAGAGTCAATCGTTTGTATGAACAATTT 1260
Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysLysLys 440
Db 1261 AAGCATATGAAGAAATGATGAACACATTCAGTGGTGGCGTAAAGTAAAGTAAAGTAAA 1320
Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCAATCAATGCAAAATATGTTAAAGGTATGAATTTACGGTTT 1365

RESULT 2

US-09-815-242-4365
; Sequence 4365, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELUTRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4365
; LENGTH: 1365
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4365

Alignment Scores: 4, 42e-242 Length: 1365
Pred. No.: 2235, 00 Matches: 447
Score: 98.30% Conservative: 3
Percent Similarity: 98.24% Mismatches: 5
Best Local Similarity: 98.28% Indels: 0
Query Match: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-815-242-4365 (1-1365)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTTGAGGCTTATCAGAACGGCTGCAGCGACGATGCAAAAAAATCGGTGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAAACTTACTAGCTGATATAAGATATATGCGTGAAGTAAAGATTAGCGTTACTT 120
QY 41 GluAlaAspValAsnPhelLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 121 GAGCGTCAGCTAACTTTAAAGTGTAAAGATTATTAACACAGTATCAGACGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTTCCGATGTAATGCAATCATTAACACAGCGCAACAGTTATTAAATAAGTTCAA 240
QY 81 AspGlnLeuThrLysLeuMetGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAATATATPCGATTAATATGTCATTAACCA 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCTACTGTTGTTATGATGTTGGTTTACAGGTCTGGTAAACACACACTGCAGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATATTGATGCGTAAAAAATACACAAAAAACCTATGTAGTTGCAGCAGATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGCTCCAGCGATTAATCAATACAAACAGTAGGGAACAAATATGATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGAGATCAAGTAAAGCCACACAAATTTGACTAATATGCTATTAACATGCT 540
QY 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACAATTTAGACTTTGATATTCATTCATACAGCAGCTCGATTACACATCGATGA 600
QY 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCAATGATGATGAATTAAGAAGTAAAGACATTCCTAAACCAACCAAAATATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGlnSerPheAspAspGln 240
Db 661 GTTGTCGATTCATAGCGGTCACAGATGCTGTCATATGTCAGAAATCITTTGACAGTCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACCTTAACCTAAATAGATGGTGATACACGTGGTGGTGCAGCT 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 761 TTATCTATTCTGTTCCGTGACACAAAAACCAATTAATTTGTTGGTATGAGTGAAGGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuLysMetGlyAsp 300
Db 841 GATGTTTACAGCTATTCATCCTGGAAGCTATGGCATCAGCTATTTCAGGCGCTGGGTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 GTGTCAAGGTTTAATTAAGAGCGCAACAGATGTGATCAAGAAAAAGCAAAAGATTTA 960
QY 321 GlnLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGln 340
Db 961 GAGAAAAAGATGCGGAGTCATCATTTACTTTAGATGATTTTATAGAACAACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAAATTTAGGACCACTGGATGATATTATAAATGATTCAGAGTATGAATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIle 380
Db 1081 AAAGGTCATGATAGCTTAATATAGTGAAGCAAAATGATCATATTAAAGCGATTATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCATGACCGCGCTGAAGAAACAATCCAGACACATTCATGATGATCAGTAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGATTGCTAAAGGTTCTGGTGCATTCATACAGAAGTCAATCGTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyGlyLysGlyLys 440
Db 1261 AACGATATGAAGAAAAATGATGAACAATTCATCGTGGCGGTAAAGGTAAAGAGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProPhe 455
Db 1321 CGCATCAATGCAAAATATGTTAAAGGTATGATTTACCGTTT 1365

RESULT 3

US-09-815-242-8030
; Sequence 8030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Onlisen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8030
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-8030
```

```
Alignment Scores:
Pred. No.: 4,44e-242 Length: 1368
Score: 2235.00 Matches: 447
Percent Similarity: 98.90% Conservative: 3
Best Local Similarity: 98.24% Mismatches: 5
Query Match: 98.28% Indels: 0
DB: 10 Gaps: 0
```

US-09-943-108a-2 (1-455) x US-09-815-242-8030 (1-1368)

```
QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCATTGAAGCTTATCGAACCGCTGCAAGCGAGATGCAAAAAATCCGTGGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGTAACTTACTGAAGCTGATATAGATATATGATGCGTGAAGTAAAGTATAGCTTACTT 120
QY 41 GluAlaAspValAlaPheLysValLysGluPheLysThrValSerGluArgAla 60
Db 121 GAGGCTGACGTAAATTTAAAGTGTAAAGAAATTTATTAACACAGTATCAGAACGCGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 TTAGGTTCCGATGTATGCAATCATTAACACCGCGGCACAAAGTATTAATATGATTTCA 240
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GATGAATTAACGAAGTTGATGGGTGGAGAAATATATCGATTATATGTCATAAACA 300
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CTTACTGTGTATGATGGTGTGTTTACAAAGTGTGTTAAACACAACTATGCGAGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCATTTATGATGCGCTAATAAATACACAAAACCTATGTTAGTTCACGACATATT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATATCAATTAACAACAGTAGGAAACAAATTCATATTCCTGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TACAGTGAAGAGATCAAGTAAAGCCACAAATTTGTAATGATATGATCAATTAACAACTGCT 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAAGAAGAACATTTAGACTTTGTAATCATTTGATACACAGCGTCGATTACATCGATGAA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 GCATTTGATGAATGAATTAAGAAGTAAAGACATTTGCTAAACCAACCAAAATTTATGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTTGTGATCAATCAAGCGGTCAAGATGCTGCAATGTTGAGAACTTTTGACGATCAA 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGATGTCACAGGTGTTACTTAATTAATTAATGATGTTGATACACGTGGTGGTCACT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
```

```
Db 781 TTATCTATTCTGTCGGTGACACAAAAACCAATAATTTGTTGGTATGAGTGAAAGTTA 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetalaSerArgIleLeuGlyMetGlyLysP 300
Db 841 GATGTTTTAGAGCTATTCCATCTGAACTATGCGATGCGATCATTTTTCAGCCCTGCTGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
Db 901 GTGTCAGGTTANTTGAAGAAGCGCAAGATGCGATCAAGAAAGCAAAAGCAAAAGTTA 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GAGAAAAAGATGCGCGAGTCACTATTTACTTTAGATGATTTTITAGACAACTTGATCAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGAAAAATTTAGGACCACTGGATGATATTAATAAATGATTCAGGTATGATAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 AAAGTCTAGATAAGCTTAATATGAGTGAAAGCAAAATTTGATCATATTAAGCGATATATC 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 CAGTCAATGACGCGCTGAAAGAAACAATCCAGACACATTTGAATGATATCAGTAAAAAG 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTAAAGTTCTGCTGCTTCATTACAAGAAAGTCAATCGTTGATGAACAATTT 1260
QY 421 AsnAspMetLysLysMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440
Db 1261 AACGATATGAGAAAGATGATGAAACAATTCATCTGCTGCGGTAAAGGTAAAGGTAAA 1320
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuProphe 455
Db 1321 CGCAATCAAAATGCAAAATATGTTAAAAAGGTATGAATTTACGGTTT 1365
```

RESULT 4

```
US-09-815-242-3890
; Sequence 3890, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 3890
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3890

Alignment Scores:
Pred. No.: 3,17e-150 Length: 1416
Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-3890 (1-1416)

QY 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ARGGTTTGGAGTTTACAAACCCGCCCTACACAGGCATGAGTAAATCCGTCGTAAG 60

QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 61 GGAAAGTTCCGAGGCGGCGPAAAGAAATGATGCGAGAAATCCGTTGCTTTATTA 120

QY 41 GluAlaAspValAsnPheLysValValLysGluPheLysThrValSerGluArgAla 60
Db 121 GAGCGGCGGTAATTTACAAAGTGGTCAAAAGATTTACAAAGCCGCTCAGAGACGGCA 180

QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 181 GTAGAGTCCGAGTATTAGAAACCTTATCACCAGCCCAACAATTTGTAATAATTTGAT 240

QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 241 GAAGAATTACGAAAGCGTTAGTTCAGAAACGGTTGAACCTGAATAATCTCCAAAATC 300

QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCGACAGTGATTATCAGACAGGTTTACAAAGGGCTGTAAACACACTTTTACTGCTAAA 360

QY 121 LeuAlaLeuLeuMetArgLysLysTyAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTAGCAAAACACTTAATGAACACTGAAACCGCTCGCTTTTAATCGCTGTCGACGTT 420

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATCGTCCAGCAGCGATTGTCAGTTCAGGTTTATAGTCAACATTAAGAGTTCGCGTT 480

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTGATATGGGAACAGATGCTAATTCAGTGGGAAATTTGTCGCAAGGGTTAGCAATGCA 540

QY 181 LysGluGluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 AAGAAGAAGAAATGATTATGCTTAATGTATGCGCGCGCGCTTTTACACATTCAGCAA 600

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleLeuMetLeu 220
Db 601 GCTTTAATGGCAATTAACAAATTAAGAGTTGGCTTAATCCCAATGAATTCGTGTA 660

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGATGCGATGACGGGCAAGATGCTGTCACGTTGCAGATAGTCTTTAATGAACAG 720

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
Db 721 CTTGGAATTTACGGGTTGTTATACCAAAATGAGCGCGCATCTCGTGGGGGGCGCG 780

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCAATTCGGGCGAGTACGCGCGCTCCGATTAATTTGTCGGTTCTCGTGAATAATTA 840

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuLysMetGlyAsp 300

```

```

; SEQ ID NO 6764
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1434)
US-09-815-242-6764

Alignment Scores:
Pred. No.: 3,23e-150 Length: 1434
Score: 1418.50 Matches: 273
Percent Similarity: 77.54% Conservative: 86
Best Local Similarity: 58.96% Mismatches: 89
Query Match: 62.38% Indels: 15
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-6764 (1-1434)

QY 1 MetalapheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCTTTTGAGAGTTTACAAACCCCTACACAGGCAATGAGTAAATCCGTCGTAAG 60
QY 21 GlyLysLeuThrGluAlaAspLysLeuMetGlnThrValGluValArgLeuAlaLeuPhe 40
Db 61 GGAAGAAGTTCCGAAGCCGAGCTAAAGAAATGATGCGAGAAATCCGTTGCGCTTTATTA 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPheLysLeuThrValSerGluArgAla 60
Db 121 GAAGCCGAGCTTATTTACAGGTCGTCAGAAATTCACAAACGCGTCAGACAGCGGCA 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLeuValGln 80
Db 181 GTAGGAGTCGAAGTATTAGAAAGCTTATCACCAGCCCAACAAATTTGTAATAATTTCTGAT 240
QY 81 AspGluLeuThrLysLeuMetGlyGluAsnThrSerLeuAsnMetSerAsnLysPro 100
Db 241 GAAGAATTAAACGAAACGTTAGTTCAGAAACGCTTGAACCTGAATAAATCTCCAAAAATC 300
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 301 CCGACAGTCATATGACAGAGGTTACAGGGGCTGGTAAACAACTTTTACTGGTAAA 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrsLysLysProMetLeuValAlaAlaAspLys 140
Db 361 TTAGCAAAACACTTAATGAAACTGAAACGCTGCTGCTGCTTTAATCGCTGCTGACCT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnLysLeuProVal 160
Db 421 TATGCTCAGCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLysValThrAsnAlaLeuLysHisAla 180
Db 481 TTGTATATGGGACAGATGCTATATCCAGTGGAAATGTTGCTCAAGGTTACGATGACA 540
QY 181 LysGluGluHisLeuAspPheValLysLeuThrAlaGlyArgLeuHisLysLeuGlu 200
Db 541 AAAGAAAAGAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLysLeuLysProAsnGluLysLeu 220
Db 601 GCTTAAATGGAGGAAATGAAACAAATTAAGAGTTGGCTAATCCCAATGAAATCTGTTA 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 661 GTTGTGATGCGATGACGGGCGAGATGCTGTCACAGTTCGAGATGTTTAAATGACAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyLysAla 260
Db 721 CTTGGAATTAATCTGGGTTGTTTATTACAAATGGACGCGGATGCTGCGGGGGGCTGCG 780
QY 261 LeuSerIleArgSerValThrGlnLysProLysPheValGlyMetSerGlnLysLeu 280

```

RESULT 6

US-09-070-927A-37

; Sequence 37, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch

; Patrick J. Dillon

; Steven Barash

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/046,655

; FILING DATE: 1997-05-16

; APPLICATION NUMBER: 60/044,031

; FILING DATE: 1997-05-06

```

; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 6729 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-070-927A-37

```

Alignment Scores:

```

Pred. No.:      3,83e-149      Length:      6729
Score:          1417.50      Matches:      273
Percent Similarity: 77.94%      Conservative: 86
Best Local Similarity: 58.96%      Mismatches: 89
Query Match:      62.34%      Indels:      15
DB:              10      Gaps:      2

```

US-09-943-108a-2 (1-455) x US-09-070-927A-37 (1-6729)

```

QY 1 MetAlaPheGluGlySerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 3722 ATGGCTTTGAGAGTTTAAACAACCGCTACAGCAGGCAATGAGTAAATAATCCGCTGAAG 3781
QY 21 GlyLysLeuThrGluAlaAspIleGlyLysMetMetArgGluValArgLeuAlaLeuPhe 40
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 3782 GGAAGAAGTTCGGAAGCCGCGTAAAGAAATGATCGGAGAAATCCGTTGGCTTTATTA 3841
QY 41 GluAlaAspValAsnPhelLysValLysGluPheIleLysThrValSerGluArgAla 60
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 3842 GAAGCCGAGCTTAATTTACAGTGGTCAAGAGATTTACAAACGCTGTCAGAGACGGCA 3901
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 3902 GTAGGAGTCGAAGTATGAGAAGCTTATCACCAGCCCAACAATTTGTAATAATTTGTTGAT 3961
QY 81 AspGluLeuThrLysLeuMetClyGlyLysThrSerIleAsnMetSerAsnLysPro 100
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 3962 GAAGAATTAACGAAAGTATGAGTTCGAGAACCGGTTGAAATGAATAATTCCTCAAAATC 4021
QY 101 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4022 CCGACAGTGATGATGACAGGCTTACAGGGGCTGTAAACAACCTTTTACTGTGTAA 4081
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4082 TTACAAACAACTTAATGAAGAACTGAAAGCGCTCGTCCGCTTTTAAATCGCTGGTGAAGT 4141
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4142 TATCGTCACAGCAGGATGATGATGAGTGTAGGTCACAAATTTAGAGTTCCCGCT 4201
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4202 TTTCATATGGGACAGATGCTAATCCAGTGAATTTGCTCGTCAAGGCTTAGCATTAAGCA 4261
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4262 AAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4321
QY 201 AlaLeuMetAsnGlnLeuLysGlnValLysGlnIleAlaLysProAsnGluIleMetLys 220
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4322 GCTTATGAGCAGGATGAAACAAATTAAGAGTGTGCTGATCCCAATGAATTTCTGTGA 4381
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240

```

```

DB 4382 GTTGTGATGCGATGACGGGCGAGATGCTGTCACAGTTGCGAGATGTTTATGAACAG 4441
QY 241 LeuAspValThrGlyValThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4442 CTTGGAATTTACTGGGGTGTATTATACAAATGGATGGCGATCTCGTGGGGGGGTGCG 4501
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4502 CTGTCAATTCGGGCGAGTAACGGGTGCTCCGATTAATTTGCGGTCTCTGTTGAAATTA 4561
QY 281 AspGlyLeuGluPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4562 ACCGATTTAGAAATTTTCCATCCCGATCGTATGCGAGTCGATCTAGTATGGGGGAC 4621
QY 301 ValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4622 ATGTTGACGCTAATTTGAAAAAGCGCAACAGATTACAGATTGAGAAAAAGCAGAACTT 4681
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4682 GCTCAAAAAATGAAAGAAACAGTTTTCGACTTTAAGGATTTTCATTCGCAATTTGGATCAA 4741
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4742 GTTATGGCATGGGACCGATGAGAGCTTATTAAAAATGATCCCTGGAATGAGTAACATG 4801
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4802 CCTGTTATGAAAAATGTCAAAAGATGATCCCAAGAGATGGGACGCAAAACGGCGATGGTC 4861
QY 381 GluSerMetThrProAlaGluArgAsnProAspThrLeuAsnValSerArgLysLys 400
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4862 TTATCAATGACCCCGAGAGACGTGAAATTCCTGATCTATTAAATCTCTAGTCGCCGCCGC 4921
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4922 AGAATTCGACGTGGTTCAGGAAATAGTGTGTTGAAGTCAATCGTATGATTAAACAATTT 4981
QY 421 AsnAspMetLysLysMetLysGlnPheThr----- 431
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 4982 AAAGAATTCCAAAAAATGATCAACAAATGTCAAAAGGGGATGATGAACATTCCTGGTATG 5041
QY 432 -----GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 446
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 5042 GATCAATGTTAGTGGCGGCGTTAAGGCAAGCTAGGTAAATG---GCCATGATCGT 5098
QY 447 MetLeuLys 449
|||||..... |||||..... |||||..... |||||..... |||||..... |||||.....
DB 5099 ATGATGAAG 5107

RESULT 7
US-09-974-300-1847
; Sequence 1847, Application US/09974300
; Patent NO. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085:500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1847
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Bacillus licheniformis

```

US-09-974-300-1847

Alignment Scores:

Pred. No.: 1,66e-148 Length: 1136
Score: 1402.00 Matches: 270
Percent Similarity: 87.30% Conservative: 60
Best Local Similarity: 71.43% Mismatches: 48
Query Match: 61.65% Indels: 0
DB: 10 Gaps: 0

US-09-943-108a-2 (1-455) x US-09-974-300-1847 (1-1136)

QY 1 MetalaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGCGATTTCGAAGATTACCGCGAGCTGCGCAACACGATCTCAAAATCCGCGGAAA 60
QY 21 GlyLysLeuThrGluAlaAspLysLysMetMetArgGluValArgLeuAlaLeuPhe 40
DB 61 GGAAGGTAACGGAGCAAGACGCTCAAGAGATGATGCGGAAGTCCGTTAGCGCTCCTT 120
QY 41 GluAlaAspValAsnPhelLysValLysGluPhelLysThrValSerGluArgAla 60
DB 121 GAACGGAGCTCAATTTAACTGCTGAAGATTTGTGAAAAGTGAAGCAAGAGCT 180
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGln 80
DB 181 CTCGGACAAGACGCTTATGAAAAGCTGACGCCGCGACAGCATCAATTAAGTGTAAA 240
QY 81 AspGluLeuThrLysMetGlyGlyGlnAsnThrSerLysAsnMetSerAsnLysPro 100
DB 241 GAAGAGTGCAGGAGCTGATGGCGCGGAGAAAGCAAGATTGCCGCTGCAAGCGCTCG 300
QY 101 ProThrValValMetValGlyLeuGlnGlyValAlaGlyLysThrThrAlaGlyLys 120
DB 301 CCGACCGCTATCATGATGCTGCTGCAGGGCGCGTAAACAGCATGTCACCGCGAAG 360
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspLe 140
DB 361 CTTGCCAACCTGCTTCGCAAAAACATAACCGCAATCGCTGCTGCGCAGCAGCATC 420
QY 141 TyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLysLeuAspLeuVal 160
DB 421 TACCGCCGCGCGCATCAACAGCTGGAACCGCTCGCAACGATCGATATCGCGGTA 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
DB 481 TTTTCATCGGTGACAGTCACTCGGTGGAGATTCACACAGCCATCGAAAGCG 540
QY 181 LysGluGlnHisLeuAspPheValLysLeuAspThrAlaGlyArgLeuHisLeuAspGlu 200
DB 541 AAAGAAGATCATCAGCATACCTTCTCATCATACGCGGAGCGCTTCATATTGATGA 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluLeuAlaLysProAsnGluLeuMetLeu 220
DB 601 GAGTGTATGATGAGTTCAGCAGTAAAGAACCGCTCAGCGGAAAGAGATTTTCCTT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAlaAsnValAlaGluSerPheAspAspGln 240
DB 661 CTCGTTGACTCCATGACGGGTCAACCGCTCAATGTTGCCAAAGCTTTAAGCAACAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyValAlaAla 260
DB 721 CTCGCTGTGACGCTGTCATCTTACGAAGCTTGACGGCGATAGAGCGGAGCGGCC 780
QY 261 LeuSerLeuArgSerValThrGlnLysProLysPheValGlyMetSerGluLysLeu 280
DB 781 TTGTCAATCCGCGCGCTGCAACACCGCGATCAATTCGCGGAATGGCGGAAAGCTT 840
QY 281 AspGlyLeuGlnLeuPheHisProGluArgMetAlaSerArgLysLeuGlyMetGlyAsp 300
DB 841 GATGGCTGGAGCGCGTTTCATCTCAACGAGTGGCTTCGCGGATATCTCGAATGGCGAT 900
QY 301 ValLeuSerLeuLeuGluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeu 320

DB 901 GTGCTACCCCTTATCGAGAAAGCCAGGCCAACGTAGACGAAGAAAGCGAAGAACTC 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
DB 961 GAGCAAAAATGAAAATATGAGCTTTACGCTGAGGACTTTTGGAAACAGCTCGGACAG 1020
QY 341 ValLysAsnLeuGlyProLeuAspLysMetLysMetLysMetLysMetLysMetLysMet 360
DB 1021 GTCCGCAATATGCGCGCTTGAGAGCTGATTAATCAATGATGCCGCGCGCAGCAAAATG 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnLysAspHisLeuLysAla 378
DB 1081 AAAGCTTAAAAAACCTGAAGCTTGATTAATAAACAGCTCAGCATATCGAAGCG 1134

RESULT 8

US-09-815-242-9156
; Sequence 9156, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9156
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1572)
US-09-815-242-9156

Alignment Scores:
Pred. No.: 1,28e-137 Length: 1572
Score: 1307.00 Matches: 253
Percent Similarity: 74.73% Conservative: 90
Best Local Similarity: 55.12% Mismatches: 102
Query Match: 57.48% Indels: 14
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-9156 (1-1572)

QY 1 MetalaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
DB 1 ATGCGATTTCGAAGATTACCGAGACGTTTTCGAAACGCTTTTAAATAATCTACGTAATAAAA 60
QY 21 GlyLysLeuThrGluAlaAspLysLysMetMetArgGluValArgLeuAlaLeuPhe 40

```

Db 1141 TCTTCGATGACCTGAAGAGCGTGAACACCGCAATTTGTTAAATCCAAGCCGTCGCGGT 1200
QY 401 ArgilealalysglySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTGCTGGTTCTCGGAATACATTTCGTGGAAGTCAATAAAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyClyLysGlyLysLysGlyLys 440
Db 1261 AACACAGGCTAAACACGCTCATCGAGSGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAAATGATGACAGCAATGGGGATTAATCCAAATACCTTCCT 1350

RESULT 9
US-09-815-242-9427
; Sequence 9427, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELIURA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/491,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9427
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1395)
US-09-815-242-9427

Alignment Scores:
Pred. No.: 1395e-137 Length: 1395
Score: 1306.00 Matches: 232
Percent Similarity: 74.73% Conservative: 91
Best Local Similarity: 54.90% Mismatches: 102
Query Match: 57.43% Indels: 14
DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-815-242-9427 (1-1395)

Qy 1 MetalAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 1 ATGGCAATTTGAAGATTAAACAGACGTTTCGACAGCGTCCTTTAAAAATCTACGTAAABAAA 60
Qy 21 GlyLysLeuThrGlnAlaAspLysLysLeuMetMetArgGluValArgLeuAlaLeuPhe 40

```

Ds 61 GGAARAATCTCTGAATCTGATGTCCAGAGGCAACCAAGAAATTCGGTTGGCTTCGCTC 120

Qy 41 GlualaaspValasnPhelysValValylsGluPheileLysThrValSerGluA-gala 60

Ds 121 GAGGCGACGCTTGCCTTGGTGTAAAGACTTTATCAAGAAAGTTTCGTCGAGCGTGCA 180

Qy 61 LeuGlySeraspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80

Ds 181 GTGGGCGATGAGTCATTCATACCTTAATCCTCGCAACAGATTATTAATAATCGTTGAT 240

Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThr-SerIleAsnMetSerAsnLysPro 100

Ds 241 GAGGAATGACCGCTTTTAGTCTTGATACCGCAAAATTTATCAAGTCACCTAGATT 300

Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaLysThrThrAlaGlyLys 120

Ds 301 CCAACCATCATGATGATGTTGGTTTACAAGGGCTGGTAAACAACCTTTTGCTGGTAAA 360

Qy 121 LeuAlaLeuLeuMetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIle 140

Ds 361 TTGGCCACAAATCTCAAGAAAGAAATGCTCGCTTTGATGATTCGCGCGGATATT 420

Qy 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160

Ds 421 TATGTCGACGCTGCCATGACCACTTAAGACCTTGGACACAGATTGATGCTGCTGTC 480

Qy 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180

Ds 481 TTTGCACTTGGACAGAGATCACCAGCTGTTGAGATTGTACGTCAAGGTTTGGACCAAGCC 540

Qy 181 LysGluGlnHisLeuAspPheValIleAspThrAlaGlyArgLeuHisIleAspGlu 200

Ds 541 CAAACTAATCATACAGCACTATGCTTGATTGATGACTCGCGGTCGTTTCAGATTGATGAG 600

Qy 201 AlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLysProAsnGlnIleMetLeu 220

Ds 601 CTCTCAAGAGAGCTTCGATGATGTAAGCAATTCGCTCAACCAATGAATCTTCCTT 660

Qy 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240

Ds 661 GTGCTTGCATGATGATGTTGGTCAGAGCAAGCAATGTCGCGTGAGTTTAATGCTCAG 720

Qy 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAla 260

Ds 721 TTGAATGACTGGGCTATCCTACCAAGATTGATGGCGATACCTCGTGGTGGTGTCT 780

Qy 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280

Ds 781 CTGCTGTCGTCACATTAAGTGAACCAATCAAGTTCACTGCTGACAGGTGAAGATT 840

Qy 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300

Ds 841 ACGGACATTGAAACCTTCCACCAGCCGATGCTAGCGGTATCCTTGGTATGGGGAT 900

Qy 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGluLysAlaLysAspLeu 320

Ds 901 ATGCTCACTTGTAGAAAGCTTCTCAGGAATGATGATGAACAAAGAGCCCTTGAATG 960

Qy 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340

Ds 961 GCTGAGAGATGCGCAACACCTTTGATTTAATGATTCATGATCAATTAGATCAG 1020

Qy 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360

Ds 1021 GTGCCAAATATGGGCGGATGAAGACTTGCTCAAGATGATTCAGGCTATGGCCAAAT 1080

Qy 361 LysGlyLysLeuAspLysLeuMetSerGluLysGlnIleAspHisIleLysAlaIle 380

Ds 1081 CCAGCACTTCAAAACATGAAGTGATGAACCGCAGATTGCTCGTAAACAGTGCATTGTG 1140

Qy 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400

Ds 1141 TCTTCGATGACCTGAAGACGTGAARACCCAGATTGTTAAATCCAGCCGTCGCCGT 1200

Qy 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPhe 420

Ds 1201 CGTATTCGCTGCTGGTCTGGAATACATTCGTCGAAGTCAATAAATTTATCAAGGACTTT 1260

Qy 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysGlyLys 440

Ds 1261 AACCAGGCTAAACAGCTCATGAGGGTGTATGCTGGG----- 1299

Qy 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454

Ds 1300 -----GATATGAATAAATGATGAGCAAAATGGGATTAATCCAATAAACCCTTCCT 1350

RESULT 10

US-09-814-041A-1

; Sequence 1, Application US/09814041A

; Patent No. US20020103104A1

GENERAL INFORMATION:

; APPLICANT: CHEEVER, CHRISTY

; APPLICANT: FECTEAU, DOUGLAS A.

; APPLICANT: LI, HU

; APPLICANT: PAYNE, DAVID J.

; APPLICANT: STEEL, ANGELA

; APPLICANT: WANG, LEI

; TITLE OF INVENTION: METHODS USING THE SRP POLYNUCLEOTIDES

; FILE OF INVENTION: AND POLYPEPTIDES COMPOUNDS MODULATING THEIR ACTIVITY

; FILE REFERENCE: GM50069

; CURRENT APPLICATION NUMBER: US/09/814,041A

; PRIOR FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: 60/191,008

; PRIOR FILING DATE: 2000-03-21

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 1569

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-814-041A-1

Alignment Scores:

Pred. No.: 1,7e-136 Length: 1569

Score: 1297.00 Matches: 251

Percent Similarity: 74.51% Conservatives: 91

Best Local Similarity: 54.68% Mismatches: 103

Query Match: 57.04% Indels: 14

DB: 10 Gaps: 2

US-09-943-108A-2 (1-455) x US-09-814-041A-1 (1-1569)

Qy 1 MetAlaPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20

Ds 1 ATGCGATTTGAAAGTTTAAACAGACGTTTCAGACAGCTCTTTTAAAAATCTACGTAAAAA 60

Qy 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40

Ds 61 GGAARAATCTCTGAATCTGATGTCCAGAGGCAACCAAGAAATTCGCTTGGCTTCGCTC 120

Qy 41 GluAlaAspValAsnPhelysValValylsGluPheileLysThrValSerGluArgAla 60

Ds 121 GAGCGCACGCTGCTGCTGCTGCTTAAAGACTTTATCAAGAAAGTTTCGTCGAGCGTGCA 180

Qy 61 LeuGlySeraspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80

Ds 181 GTGGGCGATGAGTCATTCATACCTTAATCCTCGCAACAGATTATTAATAATCGTTGAT 240

Qy 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100

Ds 241 GAGGAATGACCGCTTTTAGTCTTGATACCGCAAAATTTATCAAGTCACCTAGATT 300

Qy 101 ProThrValValMetMetValGlyLeuGlnGlyAlaLysThrThrAlaGlyLys 120

Ds 301 CCAACCATCATGATGATGTTGGTTTACAAGGGCTGGTAAACAACCTTTTGCTGGTAAA 360

```
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 361 TTGGCCAAACAACTCAGAAAGAAAGAAATGCTGCTTTGAAGATTGCGGCGAATTT 420
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 421 TATGCTCAGCTGCATTCAGCAGCTTAAGACCTTGGACACACAGATTGATGTGCTGTC 480
QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAla 180
Db 481 TTTCGACTTGGACAGAAAGTACAGCTGTGAGATTGATCAAGGTTTGAGCAAGCC 540
QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 541 CAAACTAATCATACAGCATATCTCTGATGATGATCTGCGGGTCTTGCAGATTGATGAG 600
QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
Db 601 CTCCTCATGAATGAGCTTCGTGATGTGAAGTATTGGCTCAACCAAAATGAATCTTGCT 660
QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGln 240
Db 661 GTCGTGTGATGCTATGATGTGTGAGGAGCAGCCATGTTGCGGTGAGTTTAATGCTCAG 720
QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
Db 721 TTGGAAGTACTGGGGTCACTCTACCAAGATTGATGCTGATCTACTGCTGGTGGTCTGCT 780
QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
Db 781 CTGCTGTGCTCATCATCTACATCTGGAACCAATCAAGTTCACTGATCAGGTGAAATAATT 840
QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
Db 841 ACAGATATCGAAACCTTCCACCAGACCGTATGCTAGCGATATCTTGGCATGCGGGGAT 900
QY 301 ValLeuSerLeuIleGluLysAlaGlnAspValAspGlnGlnLysAlaLysAspLeu 320
Db 901 ATGCTCATTTGATGTGAAGAGCTTCTCAGGAATACGATGACAAAGACCTTGAATG 960
QY 321 GluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGln 340
Db 961 GCTGAGAAAGATCGCGAAACACCTTTGATTTTAATGATTTTCATCGATCAATATAGTAC 1020
QY 341 ValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMet 360
Db 1021 GTGCAAAATATGGGCGCGATGGAGACTTCTCAGAGTATTCAGATATGCGCCACAAAT 1080
QY 361 LysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIle 380
Db 1081 CCAGCATTCAAAACATGAAGTGGATGAACGCCAGATTGCTGTAACCGTGCAATTG 1140
QY 381 GlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLys 400
Db 1141 TCTTCGATGACATCTGAAGACGCTGAAACCCAGATTGTTAAATCCAAACCGCTGCGCT 1200
QY 401 ArgIleAlaLysGlySerGlyArgSerLeuGlnGluValAsnArgLeuMetLysGlnPhe 420
Db 1201 CGTATTGCTGCTGGTCTGGGAATACATTCTCGGAAGTCAATAATTCATCAAGGACTTT 1260
QY 421 AsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysGlyLysLys 440
Db 1261 AACCGGCTAAACACTCATGCGAGGTGTTATGCTGGG----- 1299
QY 441 ArgAsnGlnMetGlnAsnMetLeuLysGlyMet-----AsnLeuPro 454
Db 1300 -----GATATGAATAAATGATGAAGCAAAATGGGGATTAAATCCAAATAACCTTCT 1350
RESULT 11
US-09-738-626-2255
; Sequence 2255, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
```

```
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2255
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2255

Alignment Scores:
Pred. No.: 5.84e-122 Length: 1641
Score: 1168.00 Matches: 237
Percent Similarity: 66.67% Conservative: 85
Best Local Similarity: 49.07% Mismatches: 125
Query Match: 51.36% Indels: 36
DB: 9 Gaps: 5

US-09-943-108a-2 (1-455) x US-09-738-626-2255 (1-1641)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db 4 TTGAGTCACTGTCGATCGGTGTAAGAGCGCTTCCGGCCCTCGCGCAAGGAAG 53
QY 23 LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db 64 CTCACCGGCGAGACATCAATGCAACACACACGAGATCCGCTCTCGCGCTGCTGGAAGCT 123
QY 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db 124 GAGCTTTTCATTAAACGGTTGTTCTGCTTCATTAAACCGAATCAAGGAACGCGCGTGT 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db 184 GCAGAAAGTTCTCAGGCACTCAACCCGCGCAGCAAGTCAATCAAGATCGTCAACAGGAA 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db 244 CTGTTTCAGATCCTCGGTGGCAAAACCCGCGACTGCTACTGGCCAAAACCCGCGAAC 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
Db 304 GTCATCATCTCGCAGGCTGTCAGGGTGCAGGTGAAGACCCCTCCGAGGTAACCTGTC 363
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
Db 364 AACCACTGCTCAACAGCAG---GGTCACTCTCTATGCTGTGCTGTGACCTTCAGCT 420
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyr--- 161
Db 421 CCAGGCGCAGTTACACACCTGCAATTTGGGTGAACGCGCAGGCGTTACCACTTTCGCA 480
QY 162 -----SerGluGlyAsp 165
Db 481 CCGGATCCAGGCACCACCATCGACTCCCTCGACGCAAAATGGGCACTCCCGCGGTGAT 540
```

```

Qy      166  GlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeu 185
Db      541  -----CCAGTCGAGGTAGCGCGCAGGTATCGAAGAGCCAGCGCACCCAGCAC 591

Qy      186  AspPheValIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlu 205
Db      592  GACATCGTGATCGTATGATACCGCAGGTGCGTGGTATGATGAAACCCGTGATGACTCAG 651

Qy      206  LeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSerMet 225
Db      652  GCACGCAACATCGGAGAGCCATCAACCCCTGATGAGTGTCTTTGTTCATTGACTCCATG 711

Qy      226  ThrGlnGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGly 245
Db      712  ATGTGTCAGAGCCGCTAGACACCGCGAAGCATTCGCGAGCGGTGCTGACACTTCCATCGT 771

Qy      246  ValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSer 265
Db      772  GTTGCTCTCACCAAGCTTGATGGCGAGCCGCGGTGGTGTGCTGCACTATCCATCCGTGAA 831

Qy      266  ValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeu 285
Db      832  GTACCGGCAACCCATCATCTTTGCTCCACTGTGTGTAAGAACTCGACGACTTCGACGTC 891

Qy      286  PheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIle 305
Db      892  TTCACCCAGAGCGCATCGCCAGCGAATCTTGGCGATGGGTGAGTACTGTCACTCATC 951

Qy      306  GluLysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArg 325
Db      952  GAGCAGCGCGAGCAGTGTATGATCAGGAAAGCGAGAGGTGCTGCCAGAAAGTTGGGC 1011

Qy      326  GluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
Db      1012  TCGGGGAGCTCACCTGGAGAGACTTCTTGACCAATGCTGATGATCCGCGCATGGGA 1071

Qy      346  ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
Db      1072  CCAATCGCAACATCCTCAAGATCCTCCCTGTGGTGGCAGCAGATGTCCTCAAAATGGCGG 1131

Qy      366  LeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeuGlnSerMetThrPro 385
Db      1132  ATG---GTTGATGAGAACAACATCGACCGCATCCAGGCGATATCCGCGGTATGACCCCG 1188

Qy      386  AlaGluArgAsnAspProAspThrLeuAsnValSerArgLysLysArgIleAlaLysGly 405
Db      1189  GCGAGCGGATATTCGAAGATCCTCAAGCTTCAGGCGCAAGCGCATCGCCAACCGGT 1249

Qy      406  SerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
Db      1249  TCGGGTGTACCGTGTCCGAAAGTAAACAACTTGTTGAACGCTTCTTCGAGGCTCGCAAG 1308

Qy      426  MetMetLysGlnPheThrGly----- 432
Db      1309  ATGATGGGTCAAAATGGCTGGCCAGCTTTGGCATGGGTCCTGGATCCCGCATGCAACCAAG 1368

Qy      433  ---GlyGlyLysGlyLysGlyArgAsnGlnMetGlnAsnMetLeuLysGlyMet 451
Db      1369  AACCAAGCCAAAGCGCGCAGGATAGACGGCAAGCGTAACCAAGCAAGAGGGCCCA 1429

Qy      452  AsnLeuPro 454
Db      1429  ACCCAGCCA 1437

```

RESULT 12

US-09-815-242-6166

; Sequence 6166, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

```

; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELTRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6166
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1362)
US-09-815-242-6166

```

Alignment Scores:

```

Pred. No.:      1,11e-121      Length:      1362
Score:          1164.50      Matches:      231
Percent Similarity: 71.74%      Conservative: 94
Best Local Similarity: 50.99%      Mismatches: 121
Query Match:      51.21%      Indels:      7
DB:              10      Gaps:      2
US-09-943-108A-2 (1-455) x US-09-815-242-6166 (1-1362)

```

```

Qy      3  PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
Db      4  TTGTGAAATTTAAACCGATCGTTTGTGCGCAGCGCTGCGCAATATCATGTCGCGTGGACGC 63

Qy      23  LeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
Db      64  CTCACCTGAAGACACAGTAAAGATACGCTGCGGGAAGTGCAGTGGCGCTGCTGGAGGG 123

Qy      43  AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
Db      124  GAGTGACTCTCGCGTAGTCGTGAGTTTATCAATCGCGTAAAGAGAAAGCGGTGTGT 183

Qy      63  SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGlu 82
Db      184  CATGAAGATTAAAGAGGCTCACCGCGGCGAGGAGTTGTCGCAAAATAGTCCGTACAGAA 243

Qy      83  LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
Db      244  CTGGTTGCGCGCATGGCGAGAGAACACACCCCTGACCTGGCTGGCGACCGCTGGG 303

Qy      103  ValValMetMetValGlyLeuGlnGlyLysThrThrAlaGlyLysLeuAla 122
Db      304  GTCGTACTGATGGCGGCTTGCAGAGGTGCGGTAAACACACACAGCGTTGTAACTCGGT 363

Qy      123  LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
Db      364  AAGTTCTCTCGCGGAGAGCACAGAGAGAAAGTGTGTGTGTCTGCGCAGCTTTATCGC 423

```

```
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
Db 424 CGCGCGCAATCAACAGCTGAGACCGTGGCAGGACAGGTGGCGTGTATTTCTCCCT 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisLysGlu 182
Db 484 TCTGATGTTGGTCAGAGCGCGGTAGATATCGTTAACGGCGGCTGAAAGAACCAACTG 543
QY 183 GluHisLeuAspPheValIleAspThrAlaGlyArgLeuHisLysGluAlaLeu 202
Db 544 AAATCTCAGCAGCTGCTGCTGGTGGATACCGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
QY 203 MetAsnGluLeuLysGlnValLysGluIleAlaLysProAsnGlnIleMetLeuVal 222
Db 604 ATGGACGAGATCAACAGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeu 242
Db 664 GAGCGCATGACCGCTCAGGATGGCGCAATACGCAAAAGCATTCATGAAGCGTTACCG 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
Db 724 CTTACCGCGTAGTGTGACCAAGTGGACGGGATGCCCGCGGCTGGCGGCTCTCT 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
Db 784 ATTGCTCATTATCGCAACCGATCAAGTCTCTCGTGTGCGGAGAAACTGAGCGG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
Db 844 CTGGAGCGCTTCCATCCGCGCCATCGCGTGGTGTATCTCGCATGGCGGCTACTG 903
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnGlnLysAlaLysAspLeuLys 322
Db 904 TGCGTATCGAAGATATATCAAAAGTGTGACCGCGCGAGCGCAAAATTAGCCAGC 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGlnGlnLeuAspGlnVal 341
Db 964 AAGCTGAAAAAGGTGACGCTTCGATCTCAACGACTTCTTGACGACGCTGCCCGAGG 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
Db 1024 AAAATATATGGCGCATGCTAGTCTGATGGCAAGCTGCGCGCATGGCGGAGATCCCG 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleIleGln 381
Db 1084 GATACGCTCAAGTCACAGATGGACGATAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
QY 382 SerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArg 401
Db 1144 TCGATGACGATGAAGAGCGCGCTAAGCCAGCAATCATCAAGGTTCCGCTAAACCGCT 1203
QY 402 IleAlaLysGlySerGlyArgSerLeuGlnGlnValAsnArgLeuMetLysGlnPheAsn 421
Db 1204 ATTGCTGCGGTTGCGGTATCGAGTGCAGGACGCTAACCCTCTCTCTGAAACAGTTCGAC 1263
QY 422 AspMetLysMetMetLysGlnPheThrGlyGlyLysGlyLysLysGlyLysArg 441
Db 1264 GACATCGCGCATGATGAAG-----AAAATGAAGAGCGCGGAATG 1305
QY 442 AsnGlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
Db 1306 CGAAGATGATGAGAACGATGAGGGTATGATGCCGCCCA 1344
```

RESULT 13

US-09-815-242-1949/c

; Sequence 1949, Application US/09815242

; Patent No. US20020061589A1

; GENERAL INFORMATION:

; APPLICANT: Haseibeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1949
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-1949
```

Alignment Scores:

Pred. No.:	2,4e-121	Length:	721
Score:	1158.00	Matches:	234
Percent Similarity:	99.58%	Conservative:	2
Best Local Similarity:	98.73%	Mismatches:	1
Query Match:	50.92%	Indels:	0
DB:	10	Gaps:	0

US-09-943-108a-2 (1-455) x US-09-815-242-1949 (1-721)

```
QY 1 MetalPheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLys 20
Db 713 ATGCGATTTGAGGGTTATCAGACGCTTCAAGCGAGCGATGCAAAAATCGGTGTAG 654
QY 21 GlyLysLeuThrGluAlaAspIleLysIleMetMetArgGluValArgLeuAlaLeuPhe 40
Db 653 GGTAAACTTACTGAAAGCTGATATAAGATAAATGATCGCTGAAGTAAGATTAGCGTTACTT 594
QY 41 GluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAla 60
Db 593 GAGCGTACGTAAACTTTAAAGTGTGTAAGAATTTATTAAACAGTATCAGAACCGCA 534
QY 61 LeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGln 80
Db 533 TTAGTTCCGATGTAATGCAATCATTAACACCGGCAACAGTTATTAAATAGTTCAA 474
QY 81 AspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysPro 100
Db 473 GATGAATTAACCGAGTTGCGGTGGAGAAAATACGCTGATTATATGCAATAAACCA 414
QY 101 ProThrValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLys 120
Db 413 CCTACTGTTGTTATGATGGTTGGTTTACAAGGTGCTGGTAAACAACTGCGAGTAA 354
QY 121 LeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIle 140
Db 353 TTAGCATTTATGTCGTAAATAACACAAAACCTATGTTAGTTGCGAGCATATT 294
QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProVal 160
Db 293 TATCGTCGACGAGCATTAATCAATTAACAACAGTAGGAGAAACAAATGATATTCTCTGA 234
```



```
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1144 ATATCACTCGATGACGCCGGCGAAGCGGATCGCGAAATCATCAGCGCTCGGC 1203
QY 399 LysArgileAlaLysGlySerGlyArgSerLeuGlnGluValAlaSerGluLeuMetLys 418
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1204 AAGCGCGATCGCCCTTGCTCGGTACCCAGGTGAGGCGTGGCGGCTGATCAAG 1263
QY 419 GluPheAsnAspMetLysLysMetMetLysGlnPheThrGlyGlyGlyLysLys 438
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1264 CAGCACAGCAGATCGAAGATGATGAGAAGGTCACCGCAAGGC----- 1311
QY 439 GlyLysArgAsnGlnMetGlnAsnMetLysGlyMetAsnLeuProPhe 455
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1312 -----GGCATGCCAAGATGATCGGAGGCGTGGCAGTATGTTTC 1350

RESULT 15
US-09-815-242-6868
; Sequence 6868, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Judi L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6868
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
US-09-815-242-6868

Alignment Scores:
Pred. No.: 1.09e-118 Length: 1389
Score: 1138.00 Matches: 225
Percent Similarity: 70.67% Conservative: 93
Best Local Similarity: 50.00% Mismatches: 122
Query Match: 50.04% Indels: 10
DB: 10 Gaps: 2

US-09-943-108a-2 (1-455) x US-09-815-242-6868 (1-1389)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4 TTTGAGATTATTCGATCGCTTCCAAACCTTACGTAATATCACAGGAAGGCCGT 63
QY 23 LeuThrGluAlaAspIleLysMetMetArgGluValArgLeuAlaLeuPheGluAla 42
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 64 TTACCGAAGATAATATTAAGAAACCTTACCGAAGTCGCTATGCACTTACTTGAAGCC 123
QY 43 AspValAsnPheLysValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 124 GATGTCCTTGCTGCGGTGCGTGAATTAACGCAAAAGTAAGAAAGAGCGCGTGGG 183
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValLysLysLysValGlnAspGlu 82
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 184 GAAGAAGTCAATAAAGATTAACGCCAGGCAAGAAATCTTAAAAATCGTTCAGCGTGA 243
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProThr 102
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 244 CTTGAAAAAGCCATGGGCGAAGCAATGAGCTTTAAATCTCGCAACCAACCAACCA 303
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAla 122
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 304 GTTATCTTAATGGCGGTTTACAAGGCGCGGTAAACCAACCAAGTGGTAAATGGCA 363
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleLysArg 142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 364 AAATCTTCTGCGTGAACGCCATAAAAAGAGTGTAGTGGTGTCTGCTGACGTATATCG 423
QY 143 ProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 424 CTGCTGCGNTTAACCACTTGAACCTTGGCTCAATCCGTGCGGTGATTTTTCCTCA 483
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 484 TCGATGTTAAACAAAAACCCGCTGTATATGCTAAATCGGCGCTGTGTATGCAAACTG 543
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyAsgLeuHisIleAspGluAlaLeu 202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 544 AAATCTACGATGTTGATGTGTGATGCGACGCGTCCCTACACCTTGATACAGAAATG 603
QY 203 MetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleLeuValVal 222
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 604 ATGACGAAATCAACCAAGTCCATGCTGCATTAAATCCAAATCGAAACTCTTTTCACTGT 663
QY 223 AspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAsp 242
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 664 GATGCGATGACTGTTCAAGATGCGCAAAATACAGCAAAAGCCCTTTAATGAAGCATTGCT 723
QY 243 ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSer 262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 724 CTTACAGGCTTATTTGACAAAGTGGACGGTGTATGCGCGCGGTGGTGGCGGTATCG 783
QY 263 IleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGly 282
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 784 ATTCGTCAAAATCACAGGTAAACCAATCAAAATCTTGGGTGPGGCGAGAAACAGAGGCG 843
QY 283 LeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeu 302
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 844 CTTGAGCCATTCATCTCTGATGCTGCTCCGCTATTTTGGGCATGGCGCATGTGCTT 903
QY 303 SerLeuIleGluLysAlaGlnGlnAspValAspGlnLysAlaLysAspLeuLysLys 322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 904 TCCCTTATCGAAGATCTTGAACGTTCTGTTGATCTGAAAGGCGGAAATAATGGCGGAG 963
QY 323 LysMetArgGlu---SerSerPheThrLeuAspAspPheLeuGlnLeuAspGlnVal 341
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 964 AAATTCAGAAAGCGATGATTTTACTTTAGATGATTTCCGCAACAGCTGATGCAATG 1023
QY 342 LysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLys 361
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1024 AAAAAATGGGCGCATGATGCTATGCTGAGAAATATACAGGTGCAAAAAATTTGCT 1083
QY 362 GlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIleLysAlaIleLeuGln 381
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1084 GACACGTTAAATAACAGTAGATGACAAAAGTTTGTCAAATGGAGCGATCATTAAC 1143
QY 382 SerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnValSerArgLysLysArg 401
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: February 21, 2003, 15:16:32 : Search time 22 Seconds
(without alignments)
1988.235 Million cell updates/sec

Title: US-09-943-108a-2
Perfect score: 2274
Sequence: 1 MAPEGLSERLQATMQRGK.....GKKGRNOMWLKGMNLPF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2274	100.0	455	2 H89896	hypothetical prote
2	1717.5	75.5	450	2 A11299	signal recognition
3	1708.5	75.1	450	2 A11671	signal recognition
4	1647	72.4	446	2 B47134	signal recognition
5	1584	69.7	451	2 D83960	signal recognition
6	1393	61.3	449	2 E97116	signal recognition
7	1315	57.8	518	2 H86826	signal recognition
8	1307	57.5	523	2 E98017	hypothetical prote
9	1306	57.4	523	2 F95149	signal recognition
10	1180.5	51.9	433	2 F72236	signal recognition
11	1173.5	51.6	453	2 A10833	signal recognition
12	1164.5	51.2	453	2 E65039	signal recognition
13	1164.5	51.2	453	2 A91063	signal recognition
14	1164.5	51.2	453	2 E95907	signal recognition
15	1150.5	50.6	461	2 H82306	signal recognition
16	1150	50.6	457	2 F83178	signal recognition
17	1139.5	50.1	453	2 A50400	signal recognition
18	1138	50.0	462	2 H64048	signal recognition
19	1131	49.7	456	2 F81215	signal recognition
20	1130	49.7	456	2 B1792	signal recognition
21	1122.5	49.4	482	2 S75847	signal recognition
22	1106.5	48.7	496	2 S68160	probable RNA bindi
23	1092.5	48.0	490	2 A52050	signal recognition
24	1091.5	48.0	525	2 D70747	probable fth prote
25	1077.5	47.4	550	2 T34771	signal recognition
26	1069	47.0	521	2 H87111	signal recognition
27	1066.5	46.9	452	2 G75346	signal recognition
28	1062.5	46.7	523	2 A92907	signal recognition
29	1062.5	46.7	554	2 H97682	signal recognition

30	1054	46.4	523	2 AD3280	signal recognition
31	1050.5	46.2	449	2 F90554	hypothetical prote
32	1023	45.0	564	2 S36637	signal recognition
33	1018	44.8	434	2 G82851	signal recognition
34	1009	44.4	454	2 E70448	signal recognition
35	1006	44.2	451	2 H84975	signal recognition
36	981.5	43.2	449	2 F97726	signal recognition
37	970.5	42.7	508	2 C87702	signal recognition
38	958.5	42.2	449	2 A71728	signal recognition
39	944	41.5	447	2 C54811	SRP54 protein - M
40	933.5	41.1	448	2 C82936	signal recognition
41	905.5	39.8	447	2 E70186	signal recognition
42	898	39.5	446	2 C64205	signal recognition
43	895	39.4	450	2 S73419	signal recognition
44	888	39.1	448	2 C81718	signal recognition
45	869	38.2	448	2 D71566	probable signal re

ALIGNMENTS

RESULT 1
H89896
hypothetical protein fth [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: H89896
R:kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;
ma, A.; Mizutani-Ul, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu,
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: H89896
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <KUR>
A:Cross-references: GB:BA000018; PID:g13701037; PIDN:BAB42332.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fth
C:Superfamily: signal recognition particle 54K protein

Query Match	100.0%	Score 2274;	DB 2;	Length 455;
Best Local Similarity	100.0%	Pred. No. 7.6e-109;	Mismatches 0;	Gaps 0;
Matches 455;	Conservative 0;			
QY 1	MAPEGLSERLQATMQRGKGLTEADIKIMREVELALFEADVNFVKVKEFIKTVSERA 60			
DB 1	MAPEGLSERLQATMQRGKGLTEADIKIMREVELALFEADVNFVKVKEFIKTVSERA 60			
QY 61	LGSDVMGSLTPGOQVTKIVQDELTKMGGENTINMSNKPPTVVMVGLQAGKTTAGK 120			
DB 61	LGSDVMGSLTPGOQVTKIVQDELTKMGGENTINMSNKPPTVVMVGLQAGKTTAGK 120			
QY 121	LALLMRKYNKPMVAADYRPAAINQLQTVGKQIDIPVYSEGQVKKPQIVTNALKHA 180			
DB 121	LALLMRKYNKPMVAADYRPAAINQLQTVGKQIDIPVYSEGQVKKPQIVTNALKHA 180			
QY 181	KEEHLDFVIIDTAGRLHIDEALMNELKEVKEIAKPMNEIMLVYDSMTGDAVVAESFDDQ 240			
DB 181	KEEHLDFVIIDTAGRLHIDEALMNELKEVKEIAKPMNEIMLVYDSMTGDAVVAESFDDQ 240			
QY 241	LDVTVGLTKLDGDFGGAALSIRSVTKPIKTVGMSKLDGLLELFPERWASRLIGMD 300			
DB 241	LDVTVGLTKLDGDFGGAALSIRSVTKPIKTVGMSKLDGLLELFPERWASRLIGMD 300			
QY 301	VLSLIEKAQDVQDEKAKDLKKMRSSFTLDDFLEQLDVKNLGLPDDIMKMPGNMKM 360			
DB 301	VLSLIEKAQDVQDEKAKDLKKMRSSFTLDDFLEQLDVKNLGLPDDIMKMPGNMKM 360			
QY 361	KGLDKLMSKQIDHIKAILOSMTFAERNPDPLVNSRKRIAGSGRSLOEVNRLMKQF 420			

```
Db 361 KGLDLNMXSEKQIDHKAIIQSWTPAERNPDPTLVNSKKRIAKSGRSIQEVNRLMKQF 420
Qy 421 NDMKKMMKQFTGGGKGKGRNQNMNKLGMNLPF 455
|||||
Db 421 NDMKKMMKQFTGGGKGKGRNQNMNKLGMNLPF 455

RESULT 2
Ail1299
signal recognition particle protein ffh homolog ffh [imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: Ail1299
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: Ail1299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-450 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA97145.1; PID:gl6414416; GSPDB:GN00178
A:Experimental source: strain Clp1122
C:Genetics:
A:Gene: ffh
C:Superfamily: signal recognition particle 54K protein

Query Match 75.1%; Score 1708.5; DB 2; Length 450;
Best Local Similarity 72.3%; Pred. No. 4.7e-80;
Matches 329; Conservative 65; Mismatches 56; Indels 5; Gaps 1;

Qy 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADYNEFKVKEFIKTVSERA 60
|||||
Db 1 MAFEGLAGRLQETMNRKRGKGVNEADYKEMREVRLLALFEADYNEFKVKEFIKTVSERA 60

Qy 61 LGSQVMSLTPGQOVKIVQDELTKLGGENTSIKNSKPTVTVMVGLQGAGKTTAGK 120
|||||
Db 61 VGADYKMSLTPGQOVKIVQDELTKLGGESKIGTADRPPTVIMVGLQGAGKTTAGK 120

Qy 121 LALLMKKYNKPKMLVAADYRPAALNQLQVTKGQIDIPVYSEGQVQKQIIVTNALKHA 180
|||||
Db 121 LANLKRKYNKPKMLVAADYRPAALNQLQVTKGQIDIPVYSEGQVQKQIIVTNALKHA 180

Qy 181 KEHLDFVIIDTAGRLHIDEALMNEKVEKTAKEPNEIMLVDSMTGQDANVAESFDQ 240
|||||
Db 181 KEHLDFVIIDTAGRLHIDEALMNEKVEKTAKEPNEIMLVDSMTGQDANVAESFDQ 240

Qy 241 LDVTGVTLLKLDGTRGGAALSIKRSVTOKPIKFGVSEKLDGLFHPERMASRIILGMD 300
|||||
Db 241 LEITGVVLLKLDGTRGGAALSIKRSVTOKPIKFGVSEKLDGLFHPERMASRIILGMD 300

Qy 301 VLSLIEKAQDVQDEKADLEKMKRESFTLDDFLEQIDQVKNLGLPDDIMKIPGMNKM 360
|||||
Db 301 VLSLIEKAQDVQDEKADLEKMKRESFTLDDFLEQIDQVKNLGLPDDIMKIPGMNKM 360

Qy 361 KGLDLNMXSEKQIDHKAIIQSWTPAERNPDPTLVNSKKRIAKSGRSIQEVNRLMKQF 420
|||||
Db 361 KGLDLNMXSEKQIDHKAIIQSWTPAERNPDPTLVNSKKRIAKSGRSIQEVNRLMKQF 420

Qy 421 NDMKKMMKQFTGGGKGKGRNQNMNKLGMNLPF 455
|||||
Db 421 AEMKKMMKQFTGGGKGKGRNQNMNKLGMNLPF 455

RESULT 3
Ail1671
signal recognition particle protein ffh homolog ffh [imported] - Listeria innocua (strain
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: Ail1671
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: Ail1671
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-446 <HON>
A:Cross-references: GB:D14356; NID:g439700; PIDN:BAA21691.1; PID:g2309080
A:Note: sequence extracted from NCBI backbone (NCBIN:135652, NCBI:135654)
R:Kunst, F.; Ogawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
```

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet, Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iulio, M.P.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: H69621

A:Status: preliminary; nucleic acid sequence not shown; translation not shown.

A:Molecule type: DNA

A:Residues: 1-446 <KUN>

C:Cross-References: GB:AL009126; GB:AL009126; NID:g2633902; PIDN:CAB13471.1; PID:g2633970

A:Experimental source: strain 168

C:Genetics:

A:Gene: fff

C:Superfamily: signal recognition particle 54K protein

Query Match 72.4%; Score 1647; DB 2; Length 446;
Best Local Similarity 70.1%; Pred. No. 6.2e-77;
Matches 319; Conservative 66; Mismatches 60; Indels 10; Gaps 1;

QY 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60

QY 61 LGSVMQSLTPGGQVYKIVQDELTKLMGENTSIINMSKPPVYVMYVLOGAGKTTTACK 120
DB 61 LGSVMQSLTPGGQVYKIVQDELTKLMGENTSIINMSKPPVYVMYVLOGAGKTTTACK 120

QY 121 LALLMRKYNKKEMLYAADIYRPAALINQLOTVGKQIDIPVYSEGDOVKPQOIVTNALKHA 180
DB 121 LALLMRKYNKKEMLYAADIYRPAALINQLOTVGKQIDIPVYSEGDOVKPQOIVTNALKHA 180

QY 181 KEEHLDVIIDTAGRLHDEALMNELEKVEIKAPNEIMLVDSMTGQDVNVAESFDDQ 240
DB 181 KEEHLDVIIDTAGRLHDEALMNELEKVEIKAPNEIMLVDSMTGQDVNVAESFDDQ 240

QY 241 LDVTGVVTLKDDGTGGAALSTRVTKPKFVGNSEKLDGLFELFHPERMASRILMGD 300
DB 241 LDVTGVVTLKDDGTGGAALSTRVTKPKFVGNSEKLDGLFELFHPERMASRILMGD 300

QY 301 VLSLIEKAQDWDQEKAKDLEKKMRESFTLDDFLEQLDQVKNLGLDDIMKMPGNMK 360
DB 301 VLSLIEKAQDWDQEKAKDLEKKMRESFTLDDFLEQLDQVKNLGLDDIMKMPGNMK 360

QY 361 KGLDLNMSKQIDHIKATIQSMTPAERNPDNLNVRKKRIAKGSRSLQEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHIKATIQSMTPAERNPDNLNVRKKRIAKGSRSLQEVNRLMKQF 420

QY 421 NDMMKMKQFTG-GGKKGKGRNQMNKLGNNLPF 455
DB 421 NDMMKMKQFTG-GGKKGKGRNQMNKLGNNLPF 455

RESULT 5
D83960
signal recognition particle fff [imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
C:Accession: D83960

R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83960

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-451 <STO>
A:Cross-References: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA06203.1; GSPDI A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: signal recognition particle 54K protein

Query Match 69.7%; Score 1584; DB 2; Length 451;
Best Local Similarity 67.3%; Pred. No. 1e-73;
Matches 307; Conservative 73; Mismatches 70; Indels 6; Gaps 2;

QY 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60

QY 61 LGSVMQSLTPGGQVYKIVQDELTKLMGENTSIINMSKPPVYVMYVLOGAGKTTTACK 120
DB 61 LGSVMQSLTPGGQVYKIVQDELTKLMGENTSIINMSKPPVYVMYVLOGAGKTTTACK 120

QY 121 LALLMRKYNKKEMLYAADIYRPAALINQLOTVGKQIDIPVYSEGDOVKPQOIVTNALKHA 180
DB 121 LALLMRKYNKKEMLYAADIYRPAALINQLOTVGKQIDIPVYSEGDOVKPQOIVTNALKHA 180

QY 181 KEEHLDVIIDTAGRLHDEALMNELEKVEIKAPNEIMLVDSMTGQDVNVAESFDDQ 240
DB 181 KEEHLDVIIDTAGRLHDEALMNELEKVEIKAPNEIMLVDSMTGQDVNVAESFDDQ 240

QY 241 LDVTGVVTLKDDGTGGAALSTRVTKPKFVGNSEKLDGLFELFHPERMASRILMGD 300
DB 241 LDVTGVVTLKDDGTGGAALSTRVTKPKFVGNSEKLDGLFELFHPERMASRILMGD 300

QY 301 VLSLIEKAQDWDQEKAKDLEKKMRESFTLDDFLEQLDQVKNLGLDDIMKMPGNMK 360
DB 301 VLSLIEKAQDWDQEKAKDLEKKMRESFTLDDFLEQLDQVKNLGLDDIMKMPGNMK 360

QY 361 KGLDLNMSKQIDHIKATIQSMTPAERNPDNLNVRKKRIAKGSRSLQEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHIKATIQSMTPAERNPDNLNVRKKRIAKGSRSLQEVNRLMKQF 420

QY 421 NDMMKMKQFTG-GGKKGKGRNQMNKLGNNLPF 455
DB 421 NDMMKMKQFTG-GGKKGKGRNQMNKLGNNLPF 455

RESULT 6
E97116
signal recognition particle GTPase Fff [imported] - *Clostridium acetobutylicum*
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97116

R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteriu A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97116

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-449 <KUR>

C:Cross-References: GB:AE001437; PIDN:AAK79720.1; PID:gl5024723; GSPDB:GN00168 A:Experimental source: *Clostridium acetobutylicum* ATCC824

C:Genetics:

C:Superfamily: signal recognition particle 54K protein

Query Match 61.3%; Score 1393; DB 2; Length 449;
Best Local Similarity 61.0%; Pred. No. 5.4e-64;
Matches 280; Conservative 71; Mismatches 94; Indels 14; Gaps 5;

QY 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGSLERLOATMOKMRGKGLTEADIKIMMREVRALFADVNFVKVKEFIKTVSERA 60

```
Db 1 MAFEGSLTQLOAKMKLGRGKLSKDKIDAMREYKLLALLADVAVYKYNFVYVGVGKC 60
QY 61 LGSVMQSLTPGOQVIKIVODELTKMGGENTSIINMSKPPYVMVMVGLQGAGKTTTAK 120
Db 61 LGNEVMESLTPGOQVIKIVNEELTNLMGKESKIEFEENGITVIMAVGLQGAGKTTMCK 120
QY 121 LALLMKKYNKPMVAADYRPAAINOLOTVGKOIDIPVYSEGQVKKPOQIVTNALKHA 180
Db 121 LLSLKKK-NKKPLLVACDIYRPAALKQLEVVGKSIDIVPVSMDKVPNDVSKAAMKHA 179
QY 181 KEEHDFVIIDTAGRLHIDEALMNLKEVKEIAKPNEMLVVDSMTGDAVNVVAFSDQ 240
Db 180 KENGLNVIIIDTAGRLHIDQGLMNELENKSEVNEKEILLVVDSMTGDAVNVVAFSDNK 239
QY 241 LDVTGVTILKLDGTRGGAALSTRSVTKPIYFVGMSKLDGLLELPHPERMASRILMGD 300
Db 240 LETGVVLTKLDGTRGGAALSTRMTGPKPIYVGLGEMNDIEIFHPDRMASRILMGD 299
QY 301 VLSLIEKAQDVDEKAKDLKKMRSSFTLDDLEQLDQVKNLGPLDDIMKMPGNK- 359
Db 300 VLVIEKASADEKQAKELGDRMLSQEFNFDLQAEQMKLGPGLGLEWVPGFNSS 359
QY 360 -MKGLDKLMSKQIDHIKALIOSMTPAERNPD--TLNVRKKRIAKGSRSLQEVNRL 416
Db 360 MLKGVD-LSKNEGENKKYEAIIKSMFAKERKPNLSITASTGKRRIALGSGTTVQEVNKI 418
QY 417 MKQFNDMKMKMQFTGGGKGGKGRKQNMQLKGMNLP 455
Db 419 LKNEQMKMKMKQFKGKGRKSGK-----LFGGKMPF 449

RESULT 7
H86826
signal recognition particle protein Ffh [imported] - Lactococcus lactis subsp. lactis (S
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86826
Genome Res. 11, 731-753, 2001
R:Boilotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235166; PMID:11337471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-518 <STO>
A:Cross-references: GB:AE005176; PID:g12724623; PIDN:AAK05714.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ffh
C:Superfamily: signal recognition particle 54k protein

Query Match 57.8%; Score 1315; DB 2; Length 518;
Best Local Similarity 55.4%; Pred. No. 6.1e-60;
Matches 258; Conservative 88; Mismatches 102; Indels 18; Gaps 2;

QY 1 MAFEGSLRQATQMKRGKGLTEADIKIMMREVRALFADNVKVKVEFTKTVSERA 60
Db 1 MAFENLTERLQNVFKNLGRKKIKETDVTETKEIRVALLEADVALPVVKFKTATRERA 60
QY 61 LGSVMQSLTPGOQVIKIVODELTKMGGENTSIINMSKPPYVMVMVGLQGAGKTTTAK 120
Db 61 VGHEVIDTLNPAQQLIKIVDELTAVIGSDTAETIKSPKPTIIMVVGLOGAGKTTTAK 120
QY 121 LALLMKKYNKPMVAADYRPAAINOLOTVGKOIDIPVYSEGQVKKPOQIVTNALKHA 180
Db 121 LKLVKEBNARPLMVAADYRPAALDQLKTLGQIDVFPALGTEVPAVEIVRGLSQA 180
QY 181 KEEHDFVIIDTAGRLHIDEALMNLKEVKEIAKPNEMLVVDSMTGDAVNVVAFSDQ 240
Db 181 QTHNDVILIDTAGRLQIDELLMLNEURDKALAQNEILLVDMATIGQGAANVAFENAQ 240
QY 241 LDVTGVTILKLDGTRGGAALSTRSVTKPIYFVGMSKLDGLLELPHPERMASRILMGD 300
Db 241 LEVTGVILTKLDGTRGGAALSVRHITGPKFTGTGKIDTETPHDRMSRILMGD 300
QY 301 VLSLIEKAQDVDEKAKDLKKMRSSFTLDDLEQLDQVKNLGPLDDIMKMPGNK 360
Db 301 MLFLIEKASQYDEQKALEMAEKRENTDFDNFDIDQLDQVQNMGPWEDLLMKIPGMANN 360
QY 361 KGLDKLMSKQIDHIKALIOSMTPAERNPDTLNVRKKRIAKGSRSLQEVNRLKQF 420
Db 361 PALQNMKVDKIDTAGRLHIDEALMNLKEVKEIAKPNEMLVVDSMTGDAVNVVAFSDQ 420
QY 421 NDMMKMKMQFTGGGKGGKGRKQNMQLKGM-----NLP 454
Db 421 NOAKQLAQGVMSG-----DNMKMKMQGMINPNLP 450
```

```
Db 241 LDITGVIIITKLDGTRGGAALSTIREITGPKIFTGTGSEKLTDLVEVYFDPMSRILMGD 300
QY 301 VLSLIEKAQDVDEKAKDLKKMRSSFTLDDLEQLDQVKNLGPLDDIMKMPGNK 360
Db 301 MLFLIEKAQYDEQKALEMAEKRENTDFDNFDIDQLDQVQNMGPWEDLLMKIPGMANN 360
QY 361 KGLDKLMSKQIDHIKALIOSMTPAERNPDTLNVRKKRIAKGSRSLQEVNRLKQF 420
Db 361 PGLDKVYKDPKDVARKKAVLSMTPAERHLEAELSPARRRIAGSGNSFIEVNFKIQF 420
QY 421 N-----DMKKMKMQFTGGGKGGKGRKQNMQLKGMNLP 454
Db 421 NOSKEMAGIANGDNAMQKMGAGG-----QMPNMPAGSGMP 460

RESULT 8
E98017
hypothetical protein ffh [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 17-May-2002
C:Accession: E98017
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,
Y, P.; Sun, P.M.; Winkler, M.B.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas,
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98017
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99969.1; PID:g15458797; GSPDB:GN00174
C:Genetics:
A:Gene: ffh
C:Superfamily: signal recognition particle 54k protein

Query Match 57.5%; Score 1307; DB 2; Length 523;
Best Local Similarity 55.1%; Pred. No. 1.6e-59;
Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGSLRQATQMKRGKGLTEADIKIMMREVRALFADNVKVKVEFTKTVSERA 60
Db 1 MAFESLTERLQNVFKNLGRKKISQSDVQETKRIALLLEADVALPVVKDFIKVRERA 60
QY 61 LGSVMQSLTPGOQVIKIVODELTKMGGENTSIINMSKPPYVMVMVGLQGAGKTTTAK 120
Db 61 VGHEVIDTLNPAQQLIKIVDELTAVIGSDTAETIKSPKPTIIMVVGLOGAGKTTTAK 120
QY 121 LALLMKKYNKPMVAADYRPAAINOLOTVGKOIDIPVYSEGQVKKPOQIVTNALKHA 180
Db 121 LANKLKEBNARPLMVAADYRPAALDQLKTLGQIDVFPALGTEVPAVEIVRGLSQA 180
QY 181 KEEHDFVIIDTAGRLHIDEALMNLKEVKEIAKPNEMLVVDSMTGDAVNVVAFSDQ 240
Db 181 QTHNDVILIDTAGRLQIDELLMLNEURDKALAQNEILLVDMATIGQGAANVAFENAQ 240
QY 241 LDVTGVTILKLDGTRGGAALSTRSVTKPIYFVGMSKLDGLLELPHPERMASRILMGD 300
Db 241 LEVTGVILTKLDGTRGGAALSVRHITGPKFTGTGKIDTETPHDRMSRILMGD 300
QY 301 VLSLIEKAQDVDEKAKDLKKMRSSFTLDDLEQLDQVKNLGPLDDIMKMPGNK 360
Db 301 MLFLIEKASQYDEQKALEMAEKRENTDFDNFDIDQLDQVQNMGPWEDLLMKIPGMANN 360
QY 361 KGLDKLMSKQIDHIKALIOSMTPAERNPDTLNVRKKRIAKGSRSLQEVNRLKQF 420
Db 361 PALQNMKVDKIDTAGRLHIDEALMNLKEVKEIAKPNEMLVVDSMTGDAVNVVAFSDQ 420
QY 421 NDMMKMKMQFTGGGKGGKGRKQNMQLKGM-----NLP 454
Db 421 NOAKQLAQGVMSG-----DNMKMKMQGMINPNLP 450
```

```
RESULT 9
F95149
signal recognition particle protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 17-May-2002
C:Accession: F95149
R:Uetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95149
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <KUP>
A:Cross-references: GB:AE005672; PIDN:AAK75391.1; PID:g14972772; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI287
C:Superfamily: signal recognition particle 54K protein

Query Match 57.48; Score 1306; DB 2; Length 523;
Best Local Similarity 54.99; Pred. No. 1.8e-59;
Matches 252; Conservative 91; Mismatches 102; Indels 14; Gaps 2:

QY 1 MAPEGLSERLOATMQRKRGKGLTEADIKIMREVRALFAADVNFVKVFEIKTYSER 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MAFESLTERLVNLFNRRKKGKISSEVDQEAKEIRLLEADVALPVKDFIKVRE 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 LGSVMQSLTPGQOVKIVQDELTKLMGGENTSNKSKPTVVMVGLGAGKTTTAAK 120
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 VGEHVIDTLNPAQIIKIVDELFPAVLGSDTAELIKSPKRIPTIIMVGLGAGKT 120
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 121 LALLMRKKYKPKMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQKPOQIVTNAL 180
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 LANKLKEENAPLMIRADIYRPAIDDLTKLGLQDIPVPAFGTEVPFAVEIRQGLE 180
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 181 KEELHDFVIIDTAGRLHIDALMNEKEVKEIAKPNEMLVVDSMTGQDANVAESF 240
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 QTNENYVLIDTAGRLHIDALMNEKEVKEIAKPNEMLVVDSMTGQDANVAESF 240
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 241 LDVTGVTILKLDGTRGGAALSIRSVTQKPIKFGVMSEKLDGLFELFHPERMA 300
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 LEVTGVTILKLDGTRGGAALSIRSVTQKPIKFGVMSEKLDGLFELFHPERMA 300
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 301 VLSLIEKAQDVQDEKAKDLKKMRESSFTLDDFLEQLDQVKNLGLDIDMKMIP 360
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 MLTIEKASQETDEKALEMAEKRENTFDFIDQLDQVKNLGLDIDMKMIP 360
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 361 KGLDKLMSKQIDHKAIIQSMTPAERNPDPLNVRKRIAKSGRSLSQEVNRLMK 420
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 361 PALQNMKVDROQIAKRAIVSSMTPEERNPDPLNVRKRIAKSGRSLSQEVNRLMK 420
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 421 NDMKKMKQFTGGKKGKGNQNMMLKGM-----NLP 454
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 421 NQAKQLMQGVSG-----DNMKWKQMGINPNKLP 450
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
F72236
signal recognition particle protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72236
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
```

```
A:Accession: F72236
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <ARN>
A:Cross-references: GB:AE001802; GB:AE000512; NID:g4982133; PIDN:AAD36632.1; PID:g4
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1565
C:Superfamily: signal recognition particle 54K protein

Query Match 51.9%; Score 1180.5; DB 2; Length 433;
Best Local Similarity 53.4%; Pred. No. 3.4e-53;
Matches 234; Conservative 92; Mismatches 99; Indels 13; Gaps 5;

QY 3 FEGLSERLOATMQRKRGKGLTEADIKIMREVRALFAADVNFVKVFEIKTYSER 62
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 FENIQEKLRSYFKNLGSRGKITEKNKDAIREVKLSLEADVNFVKVFEVDVHLQALG 61
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 63 SDVMQSLTPGQOVKIVQDELTKLMGGENTSNKSKPTVVMVGLGAGKTTTAAK 122
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 EEVLSLTPDQOQFIKIVDELVRINGEKNEPLRVHR-PAPIMVGLGSGKTTTCAK 120
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 123 LLMRKKYKPKMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQKPOQIVTNAL 182
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 KLL-KKEGNEPLLYAADLYRPAVDOLVKLGQVGVNVHDYNT-PVEIVKEAYDVA 178
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 183 EHLDFVIIDTAGRLHIDALMNEKEVKEIAKPNEMLVVDSMTGQDANVAESF 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 TGKDVILVDTAGRLHIDALMNEKEVKEIAKPNEMLVVDSMTGQDANVAESF 238
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 243 VTGVTILKLDGTRGGAALSIRSVTQKPIKFGVMSEKLDGLFELFHPERMA 302
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 239 LTFGVNMDGDAKGIVLSIKYVTKGPKVFGTSEKLDGLFELFHPERMA 298
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 303 SLIEKAQDVQDEKAKDLKKMRESSFTLDDFLEQLDQVKNLGLDIDMKMIP 362
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 299 SLIEKVERELDEKMKKSAEFLKAEFTLEDPEQLQEMKKLGLSILEMPLGAP 356
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 363 LDKLNSEKQIDHKAIIQSMTPAERNPDPLNVRKRIAKSGRSLSQEVNRLMK 422
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 357 --DVENSEKELKKEIATINSMTTEERNPGIINASKRRIARGSTTVQDVNKL 414
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 423 MKMKMKQFTGGKKGKGNQNMMLKGM-----NLP 454
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 415 MKALMRM-----KKGR 426
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
AI0833
signal recognition particle protein [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AI0833
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Chur
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Pa
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Ouall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica
A:Reference number: AB0502; PMID:11677608
A:Accession: AI0833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05856.1; PID:g16503831; GSPDB:GN00176
C:Genetics:
A:Gene: STY2864
C:Superfamily: signal recognition particle 54K protein

Query Match 51.6%; Score 1173.5; DB 2; Length 453;
Best Local Similarity 51.4%; Pred. No. 8.2e-53;
Matches 233; Conservative 93; Mismatches 120; Indels 7; Gaps 2;
```

```
QY 3 FEGLSERLAQTMQMRGKGLTEADIKIMREVRLALFEADVNFKVVEFIKTSYERALG 62
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 2 FDNLTDLRSRLRNISGRGLTENVKTLREVRLALLEADVALPVVREFINRVKEKAVG 61
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 63 SDVMOSLTPGQGVKIVQDELTKLMGGENTSINMSKPPVVMVGLQAGKTTTAGKLA 122
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 62 HEVNSLTPGQGEFVKIVRSELVAAMGEENQTLNLAAQPPAVVLMAGLQAGKTTSVGKLG 121
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 123 LLMRKKYNKPMVAADIVRPAAINQLOTVGKQIDIPVYSEGQVQKPOQIVTNALKHAKE 182
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 122 KFLREKHKKKVYVSAADVYRPAAIKQLETAEQVGVDFPDSVQGVKPDIVNAAKKEAKL 181
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 183 EHLDFVIIDTAGRLHIDEALMNELKEVEIAKPNEIMLVDSMTGQDAVNVAESFDQOLD 242
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 182 KEYDVLVDTAGRLHVDDEAMDEIKQVHASINPVETLFFVDMATGQDAANTAKAFNEALP 241
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 243 VTGVTLTKLGDTRGGAALSIRSVTKPKIVGMSKGLDGLFLFHPERMASRIILGMDVL 302
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 242 LTGVVLTAKVDGDARGGAALSIRHITGPKIKPLGVGKTEALEPFPHPDRIASRIILGMDVL 301
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 303 SLIEKAQDVDOEKADLEKKMRE-SSFTLDDLEQLDQVKNLGPLDDIMKMIKGMKMK 361
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 302 SLIEDIESKVDRAQAELKATLKKGDGFDLNDLEQLQKMKNGGWSLWGLKPLFGMGQIP 361
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 362 GDLKLNKSEKQIDHIIKAIQSMTPAERNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQFN 421
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 362 DNKSKQMDKVLYRMEALINSMTKERAKPELIKGSKRKRRIAGCGQVQDVNRLKQFD 421
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 422 DNKMKMKQFTGGGKGGKGRNQMNKLGMINLP 454
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 422 DMQRMMK-----KMKKGGAAMKMSRGMMP 448
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|

RESULT 12
A:1063
signal recognition particle protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
R:Blattner, R.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65039
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-453 <BLAT>
A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AAC75659.1;
A:Experimental source: strain K-12, substrain MG1655
R:Bystroem, A.S.; Hjalmarsson, K.J.; Wikstroem, P.M.; Bjorker, G.R.
EMBO J. 2, 899-905, 1983
A:Title: The nucleotide sequence of an Escherichia coli operon containing genes for the
A:Reference number: A30380; MUID:84057772; PMID:6357787
A:Accession: S07178
A:Molecule type: DNA
A:Residues: 86-453 <BYS>
A:Cross-references: EMBL:X01818; NID:g43141; PIDN:CAA25957.1; PID:g43142
C:Genetics:
A:Gene: ffh
A:Map position: 57 min
C:Superfamily: signal recognition particle 54k protein

Query Match 51.2%; Score 1164.5; DB 2; Length 453;
Best Local Similarity 51.0%; Pred. No. 2.4e-52;
Matches 231; Conservative 94; Mismatches 121; Indels 7; Gaps 2;

QY 3 FEGLSERLAQTMQMRGKGLTEADIKIMREVRLALFEADVNFKVVEFIKTSYERALG 62
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 2 FDNLTDLRSRLRNISGRGLTENVKTLREVRLALLEADVALPVVREFINRVKEKAVG 61
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 63 SDVMOSLTPGQGVKIVQDELTKLMGGENTSINMSKPPVVMVGLQAGKTTTAGKLA 122
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 62 HEVNSLTPGQGEFVKIVRSELVAAMGEENQTLNLAAQPPAVVLMAGLQAGKTTSVGKLG 121
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 123 LLMRKKYNKPMVAADIVRPAAINQLOTVGKQIDIPVYSEGQVQKPOQIVTNALKHAKE 182
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 122 KFLREKHKKKVYVSAADVYRPAAIKQLETAEQVGVDFPDSVQGVKPDIVNAAKKEAKL 181
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 183 EHLDFVIIDTAGRLHIDEALMNELKEVEIAKPNEIMLVDSMTGQDAVNVAESFDQOLD 242
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 182 KEYDVLVDTAGRLHVDDEAMDEIKQVHASINPVETLFFVDMATGQDAANTAKAFNEALP 241
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 243 VTGVTLTKLGDTRGGAALSIRSVTKPKIVGMSKGLDGLFLFHPERMASRIILGMDVL 302
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 242 LTGVVLTAKVDGDARGGAALSIRHITGPKIKPLGVGKTEALEPFPHPDRIASRIILGMDVL 301
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 303 SLIEKAQDVDOEKADLEKKMRE-SSFTLDDLEQLDQVKNLGPLDDIMKMIKGMKMK 361
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 302 SLIEDIESKVDRAQAELKATLKKGDGFDLNDLEQLQKMKNGGWSLWGLKPLFGMGQIP 361
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 362 GDLKLNKSEKQIDHIIKAIQSMTPAERNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQFN 421
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 362 DNKSKQMDKVLYRMEALINSMTKERAKPELIKGSKRKRRIAGCGQVQDVNRLKQFD 421
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
QY 422 DNKMKMKQFTGGGKGGKGRNQMNKLGMINLP 454
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 422 DMQRMMK-----KMKKGGAAMKMSRGMMP 448
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
```

signal recognition particle protein VC0560 [imported] - Vibrio cholerae (strain N18)

C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

R:Accession: H82306
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson,
chardson, D.; Ermlolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellescu,
l, R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.W.
Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A8035; MUID:20406833; PMID:10952301
A:Accession: H82306
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <HEI>
A:Cross-references: GB:AEO04142; GB:AEO03852; NID:g9654988; PIDN:AAF93728.1; GSFPD:
A:Experimental source: serogroup O1; strain NI6961; biotype El Tor
C:Genetics:
A:Gene: VC0560
A:Map position: 1
C:Superfamily: signal recognition particle 54K protein

Query Match 50.58; Score 1150.5; DB 2; Length 461;
Best Local Similarity 50.88; Pred.No. 1.2e-51;
Matches 228; Conservative 91; Mismatches 127; Indels 3; Gaps 2;

QY 3 FEGLSERLQATMKMGKGKLTADIKIMREVRALFEADVNEKVVKERIKTVSERALG 62
|| : || | : : : ||| ||| : || : ||| ||| ||| : ||| : ||| :
Db 2 FNELDLRLSKTLANISGRULETNIKDITREVMALLEADVLPVVDVFNRYKEGAVG 61

QY 63 SDVGQSITPGQCQVIKTQDBETKLGMGNSTINSNKSPPTVVMMGVGLQAGKGTAKGLA 122
|| : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 62 VEYSKSITPQQCFIKIVRQQLFVAENGSENALAAQPFAVLMAGLQGAGKTSYGKLS 121

QY 123 LLMRKKYNKPLMVLAADIYRPAAINLOLTPIYSGDGQVXPOOILVTNLKHAKH 182
| : : : : || : || : ||| ||| : ||| : ||| : ||| : ||| : ||| :
Db 122 KLLKERDKKVLVVSADYVPAPAKIQLETLANDLGVDFFPPQCKPIDIANRADHAKK 181

QY 183 EHLDFVIDTAGRLHIDEALMNELKYVEIAKPNEIMLVDSMTGDQDVNVAESDDQLD 242
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 182 KYFDVLIVDTAGRLADEQMNAEIQALHFKAINEPVETLFVVDAMTGDAADAKAFQDALP 241

QY 243 VTGYTLAKLDGDRGGGAALSIRSVTKPKIFCVGSEKLDGLEPHPERMASRTLIMGDVL 302
:||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 242 LFGVFLTKVDGDARGAALUSRVHTIGPKPIFGVGKTDALEPFHPORIASTRILGMDVL 301

QY 303 SLIEKAQDYVDQBKADLEKKMRE--SSFTLLDDFLEQLDQVKNLGFDDIMKMTPGNMKHK 361
|| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 302 SLIEDLRNYVDQBKAEKLAKFKFKKGFDEDFEQLGOMKNMGMMLDKLDPGMSQLP 361

QY 362 GLDLNMSEKOIDIHIIKAITOSTWPAERNNDPLLNSRKRIATAKGSRSLOEVNLMKFQN 421
|| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 362 ADMKDQVDYRFVFQFMALINSTWTKEQRNPDLIKGSRKKARIAGSQTQODVYRNLLKQFT 421

QY 422 DKMKRMKQFTGGGKKKKRNOQMNLKG 450

Db 422 QMKKMKMKMQQG--GMKGMNRNMQGLMGG 448

Search completed: February 21, 2003, 15:19:26
Job time : 24 secs

[illegible]

RESULT 15
H82306

GenCore version 5.1.3

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:13:26 ; Search time 14 Seconds

(without alignments)
1347.981 Million cell updates/sec

Title: US-09-943-108A-2

Perfect score: 2274

Sequence: 1 MAFGLSERLQATQMKMRGK.....GKGRNQMWLKGMLPFF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	72.4	446	1 SR54_BACSU	P37105 bacillus su
2	1286	56.6	516	1 SR54_STRMU	Q34431 streptococc
3	1164.5	51.2	453	1 SR54_ECOLI	P07019 escherichia
4	1138	50.0	462	1 SR54_HARIN	P44518 haemophilus
5	1122.5	49.4	482	1 SR54_SYNY3	P74214 synchocyst
6	1106.5	48.7	485	1 SR54_SYNP7	Q53311 synchococc
7	1091.5	48.0	521	1 SR54_MYCTU	Q33013 mycobacteri
8	1069	47.0	521	1 SR54_MYCLE	P37107 mycobacteri
9	1023	45.0	564	1 SR54_ARATH	O67615 aquifex aso
10	1009	44.4	454	1 SR54_AQUAE	P57473 bacterera ap
11	1006	44.2	451	1 SR54_BUCAI	O07347 thermus aqu
12	1005.5	44.2	429	1 SR54_THBAQ	Q92d20 rickettsia
13	958.5	42.2	449	1 SR54_RICPR	Q01442 mycoplasma
14	944	41.5	447	1 SR54_MYCMY	P47294 mycoplasma
15	898	39.5	446	1 SR54_MYCGE	P75054 mycoplasma
16	895	39.4	450	1 SR54_MVCPN	Q92k62 helicobacte
17	854	37.6	448	1 SR54_HELPJ	P56005 methanococc
18	848	37.3	448	1 SR54_HELPJ	Q57565 methanococc
19	787	34.6	451	1 SR54_METJA	O07853 sulfolobus
20	761.5	33.5	446	1 SR54_SULAC	O972e7 sulfolobus
21	759.5	33.4	443	1 SR54_PYRAB	O9vies pyrococcus
22	748.5	32.9	443	1 SR54_PYRHO	O59307 pyrococcus
23	723.5	31.8	451	1 SR54_ACIAM	P70722 acidianus a
24	721	31.7	442	1 SR54_METTH	O27376 methanobact
25	701.5	30.8	433	1 SR54_ARCFU	Q39633 archaeoglob
26	695.5	30.6	447	1 SR54_SULSO	Q972e7 sulfolobus
27	693	30.5	456	1 SR54_HALN1	O9hmn5 halobacteri
28	680	29.9	450	1 SR54_THEAC	O9hkt0 thermoplasm
29	667	29.3	455	1 SR54_THEVO	Q97978 thermoplasm
30	664.5	29.2	441	1 SR54_AERPE	Q9yb62 aeropyrum p
31	630.5	27.7	495	1 SR53_ARATH	P49967 arabidopsis
32	628.5	27.6	499	1 SR52_LYCES	P49972 lycopersico
33	628.5	27.6	541	1 SR54_YEAST	P20424 saccharomyc

ALIGNMENTS

RESULT 1
SR54_BACSU STANDARD; PRF; 446 AA.
AC P37105, 27.1 496 1 SR51_LYCES
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN PFH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=933398695; PubMed=8335643;
RT Honda K., Nakamura K., Nishiguchi M., Yamane K.;
RT "Cloning and characterization of a Bacillus subtilis gene encoding a
RT homolog of the 54-kilodalton subunit of mammalian signal recognition
RT particle and Escherichia coli Ffh."
RL J. Bacteriol. 175:4885-4894(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsche M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandebol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).

```

RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94197732; PubMed=7511896;
RA Nakamura K., Nishiguchi M., Honda K., Yamane K.;
RT "The Bacillus subtilis SRP54 homologue, Ffh, has an intrinsic GTPase
RT activity and forms a ribonucleoprotein complex with small cytoplasmic
RT RNA in vivo.";
RL Biochem. Biophys. Res. Commun. 199;1394-1399(1994).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES.
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE N-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D14356; BAA21691.1; -.
CC EMBL; Z99112; CAB13471.1; -.
CC PIR; B47154; B47154.
CC HSP; 007347; 1FFH.
CC Subtilist; BG10830; ffh.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SPB.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02881; SRP54_N; 1.
CC Pfam; PF02978; SRP_SPB; 1.
CC ProDom; PD000819; SRP54; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMs; TIGR00959; 3a05010s1; 1.
CC DR SIGNAL; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding; Membrane;
KW Complete proteome.
FT DOMAIN 1 296 G-DOMAIN (BY SIMILARITY).
FT DOMAIN 297 446 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 108 115 GTP (BY SIMILARITY).
FT NP_BIND 191 195 GTP (BY SIMILARITY).
FT NP_BIND 249 252 GTP (BY SIMILARITY).
SQ SEQUENCE 446 AA; 49541 MW; D71943B23374633D CRC64;

Query Match 72.4%; Score 1647; DB 1; Length 446;
Best Local Similarity 70.1%; Pred. No. 2,le-78;
Matches 319; Conservative 66; Mismatches 60; Indels 10; Gaps 1;

QY 1 MAFEGLSERLOATMOMRGKLTPEADIKIMREVRALFLPADYNFKVKKEFKTVYSERA 60
DB 1 MAFEGSLADRLQQTISKRGKGVSEQDVKEMREVRALLADYNFKVKVDFKVKYSERA 60
QY 61 LGSVDVMSLTTPGOQIVQIVQDELKLGNGENTSNKSNKPTVYVMYVLOGAGKTTTACK 120
DB 61 VGOVVKSTTPGOQIVQIVQDELKLGNGENTSNKSNKPTVYVMYVLOGAGKTTTSGK 120
QY 121 LALLMRKKYKPKMLVAADYVRAAINQLQGVKIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LANLRRKKNRPMVAADYVRAAIKQLTIGKQLDMVPSLGDQVSPVEIAKQALEKA 180
QY 181 KEEHLDVIDTAGRLHIDELMELKEVKEIAKPNELMYVDSMTGDVANNVAESFDDQ 240
DB 181 KEEDYDVIDTAGRLHIDELMELKEVKEIAKPNELMYVDSMTGDVANNVAESFDDQ 240
QY 241 LDVTGVTFLNKDGRGAALISRSVQKPIKFGVGNSEKLDGLFLFPHRMASRLNGMD 300

```

```

Db 241 LGTGVVTLKLDGTRGGAALSIRAVTNTPIKPAIGLGEKLDALFPHFPMASRLNGMD 300
QY 301 YLSIEKAQQVDQVEKAKOLEKKMRESSFTLDDFLEQLDQVKNLGLDDIMKMPGNKM 360
Db 301 VLTIEKAQASVEDKAKLEQKMETMSFTLDDFLEQLDQVKNLGLDDIMKMPGNKM 360
QY 361 KGLDKLNNSEKQDHIKAIQSMTPAERNNDPTLVNRRKRIAKSGRSLOEYVRLNKKQF 420
Db 361 KGLKNIQVDEKQLNHEVALIKSMTVLEKEQDPTIINARRARIKAGSGTSVGEVARNLLKQF 420
QY 421 NDKKMMKQFTGGCKGKGRKNQNMMLKGMNLPF 455
Db 421 DEMKKMMKQNTNMSKGGK-----KGFKLPF 445

RESULT 2
SR54_STRMU
ID SR54_STRMU STANDARD; PRT; 516 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN FFH.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH1005;
RA Gutierrez J.A., Cvitkovitch D.G., Brady L.J., Hamilton I.R.,
RA Hillman J.D., Blewels A.S.;
RT "Ffh of Streptococcus mutans is involved in acidurance.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=JH1005;
RX MEDLINE=96312359; PubMed=8763945;
RA Gutierrez J.A., Crowley P.J., Brown D.P., Hillman J.D., Youngman P.,
RA Blewels A.S.;
RT "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements.";
RT J. Bacteriol. 178:4166-4175(1996).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES.
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH.
CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U85882; AAB48050.1; -.
CC EMBL; U85883; AAC44500.1; -.
CC HSP; 007347; 2FFH.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SPB.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02681; SRP54_N; 1.

```

```
DR Pfam: PF02978; SRP_SPB; 1.
DR ProDom: PD000819; SRP54; 1.
DR TIGRFAMs: TIGR00959; 3a0501s01; 1.
DR PROSITE: PS00300; SRP54; 1.
DR Signal recognition particle; GTP-binding; RNA-binding.
FW DOMAIN 1 296 G-DOMAIN (BY SIMILARITY).
FT DOMAIN 297 516 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 108 115 GTP (BY SIMILARITY).
FT NP_BIND 191 195 GTP (BY SIMILARITY).
FT NP_BIND 249 252 GTP (BY SIMILARITY).
FT CONFLICT 49 V -> A (IN REF. 2).
SQ SEQUENCE 516 AA; 56968 MW; E025D3124C8D6CE1 CRC64;

Query Match 56.6%; Score 1286; DB 1; Length 516;
Best Local Similarity 55.1%; Pred. No. 1e-59;
Matches 254; Conservative 82; Mismatches 115; Indels 10; Gaps 2;

QY 1 MAFEGLSERLQATQMKRGKGLTEADIKIMRREVRALALFADYVFKVKEFKTVSERA 60
D 1 MAFESLTERLQGVFNLRGKRLSEKDVQEVTKERIALLEADVLPVVKKEFKRYKRA 60
QY 61 LGSVNMOSTPQQQVIKIVQDELTKLMGNGENSINMSKPTVTVMVGLQAGKTTTACK 120
D 1 VCHEVIDTLDPSSQIKIKNVEELTAVLGSETAEIEKSKIPTIIMVYGLQAGKTTTACK 120
QY 121 LALLKKYKKPKMLVAADIVYPAAINOLQTVGKQIDIPVYSEGDOVKPQOIVTNALKBA 180
D 1 LANKLKEENRPLMIADIVYPAIDOLKILGQQINPVFDMGTEHSANVEIVSQGLAQ 180
QY 181 KEHLDFVATIDRAGRLHIDEALMNEKKEVEIAKPNEMLVVDSMTQDAVNVAEFDDQ 240
D 1 KENRDYVLIDRAGRLQIDELKMLPELDRDKALANPEILLVYDSMIGEEANVAREFNOQ 240
QY 241 LDVTGTLTKLGDTRGGAALSISVYQKPIKPVGMSEKLDGLELFPFPERMAGRIIGMD 300
D 241 LEVIGVILTKIDGTRGGAALSVRQITGPKIKPTGTGKITDIETFFPDRMSSRIIGMD 300
QY 301 VLSLIEKAQDDVDKADLEKKMRESSFTLDDFOLDQVKNLGLPDLTKMIPGMNKM 360
D 301 LLFLIEKASQYDEKSELEAKRENSFDFNDFIEQLDQVNGSMEDILKIPGMANN 360
QY 361 KGLDKLNMSEKQIDHIKAIIOISMTPAERNPNPDLNYSRKKRIAGSGRSLEQVNRMLKOF 420
D 361 PALANVEVDEGEIARKIAVSSMPPEERNPDLLTPSRRRRIASGSGTFFVNNKEIKDF 420
QY 421 NDMKKMKMKQFTGGGKGGKGR-----NQMQN-----MLKGM 451
D 421 NQAKRMQGVMSGDKNYKMQNGINPNMNMNNDSSALEGM 461

RESULT 3
SR54_ECOLI STANDARD; PRT; 453 AA.
ID SR54_ECOLI
AC P07019; P77007;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog) (P48).
GN F5H OR B2610 OR E3904 OR ECS3473.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN SEQUENCE FROM N.A.
RX MEDLINE=84057772; PubMed=6357787;
RA Bystrom A.S., Hjalmarsson K.J., Wikstrom P.M., Bjoerk G.R.;
RT "The nucleotide sequence of an Escherichia coli operon containing
RT genes for the tRNA(mIG)methyltransferase, the ribosomal proteins S16
RT and L19 and a 21-K polypeptide.";
RL EMBO J. 2:899-905(1983).
RN [2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshina T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP FUNCTION.
RX MEDLINE=91029490; PubMed=2171778;
RA Ribes V., Roemisch K., Giner A., Dobberstein B., Tollervey D.;
RT "E. coli 4.5S RNA is part of a ribonucleoprotein particle that has
RT properties related to signal recognition particle.";
RL Cell 63:591-600(1990).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE=93063265; PubMed=1279430;
RA Lührink J., High S., Wood H., Giner A., Tollervey D., Dobberstein B.;
RT "Signal-sequence recognition by an Escherichia coli ribonucleoprotein
RT complex.";
RL Nature 359:741-743(1992).
RN [8]
RP CHARACTERIZATION.
RX MEDLINE=93063266; PubMed=1331806;
RA Phillips G.J., Silhavy T.J.;
RT "The E. coli ffh gene is necessary for viability and efficient
RT protein export.";
RL Nature 359:744-746(1992).
RN [9]
RP STRUCTURE BY NMR OF 410-434.
RX MEDLINE=97053672; PubMed=8898086;
RA Oh D.-B., Yi G.-S., Chi S.-W., Kim H.;
RT "Structure of a methionine-rich segment of Escherichia coli Ffh
RT protein.";
RN [9]
```



```
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome. 295 G-DOMAIN.
FT DOMAIN 296 453 M-DOMAIN.
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 190 194 GTP (BY SIMILARITY).
FT NP_BIND 248 251 GTP (BY SIMILARITY).
SQ SEQUENCE 462 AA; 50843 MW; 16AC32089A158F7A CRC64;

Query Match 50.0%; Score 1138; DB 1; Length 462;
Best Local Similarity 50.0%; Pred. No. 3.9e-52;
Matches 225; Conservative 93; Mismatches 122; Indels 10; Gaps 2;

QY 3 FEGLSERLQATMOKMRGKGLTEADIKIMMRVRLALFEADVNFVKVKEIKTVSRALG 62
DB 2 FENLSRLSKLTUNITGKGRUTENIKETLREVRMALLEADVLPVVRFEIAKVRKESALG 61

QY 63 SDYMSLTGQGVKIVQDELTKLMGSENTSINMSNKPPTVMVMVGLQGAGKTTAGKLA 122
DB 62 EEVNKSITGQBELKIVQRELEKAMGEANESLNLATQPPAVILMAGLOGAGKTTSGKLA 121

QY 123 LLMRKYNKKPMLVADIRPAALNQLQTVGKQIDIPVYSGDQVKKPQIVTNALKHAK 182
DB 122 KFLREHKKVLLVWSADYRPAATKOLETLAQSGVDFPDDVQKQNPVDIAKALADAKL 181

QY 183 EHLDFVITDAGRLHIDEALMNELEKVEKEIAKPNEIMLVDSMTGQADVNVASFDDQLD 242
DB 182 KFDVLIVDTAGRLVDETMDELKQVHAALNPITFLTVDAMTGDQRAANTAFKAFNEALP 241

QY 243 VTGWTLLKDGSDTRGGAALSIRSVTQPKIFVGMSEKLDGLFHPHMASRILGMDVL 302
DB 242 LTGVILTVDGARGAALSIRQITGPKIFLGVGKTEALEPFPDRVASRILGMDVL 301

QY 303 SLIEKAQDVQEKAKLEKMRP-SFLLDFLEQLQDVKNLPLDIDMKMIPGNKK 361
DB 302 SLIEDLERSVDEKAEKMAQFKKGDFTLDDFRELQIEMKKGMMKMLEKLPFGAKNLS 361

QY 362 GLDKLNMSKQDIDHKAIIQSMTPAESNNPDILNYSRKKRIAKSGRSLOEVRNLKQFN 421
DB 362 EHVKNQVDKMFVRMEALINSWTLKERANPDIIKGSRRRIALSGTGVQDVNKLKQPD 421

QY 422 DMKMMKQFTGGGKGGKGRNQMONMLKGM 451
DB 422 EMQRMKKVRKGS-----MAKMMRGM 442

RESULT 5
SR54 SYNY3
ID SR54.SYNY3 STANDARD; PRT; 482 AA.
AC P74214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN F5H OR S1R1531.
OS Synechocystis sp. (strain POC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain POC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
```



```
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74697; CAA98978.1; -.
DR EMBL; AE007120; AAK47310.1; -.
DR HSSP; O07347; 2FFH.
DR TIGR; MT2984; -.
DR Tuberculist; Rv2916c; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 301 G-DOMAIN (BY SIMILARITY).
FT NP_BIND 302 525 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 196 200 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SQ SEQUENCE 525 AA; 55001 MW; 36F6227BC367817C CRC64;

Query Match 48.0%; Score 1091.5; DB 1; Length 525;
Best Local Similarity 47.0%; Pred. No. 1.1e-49;
Matches 224; Conservative 88; Mismatches 132; Indels 33; Gaps 6;

QY 3 FEGLSERLATQWKRGKGTETADIKIMRVRVRLALFADVNFVYKFEIKTVSERALG 62
DQ ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| :
DQ 2 FESLSRLTAALQGLRGRLTDADATTREIRLALLADVSLVPVAFTHRIKERARG 61
QY 63 SDVMQSTPGQGVKTVDETKLMGGENTSINXNKPETVMYVGLGAGKTTTAGKLA 122
DQ : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 62 AEVSSALNPAQGVKIVNEELISLIGGETRELAFATPTPTVVMLAGLQSGKTTLAGKLA 121
QY 123 LLMRKYNKPKMLVRAADYRPAALNQLQTVGKQIDIPVYS-----EGDQVRPQQIVTN 175
DQ : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 122 ARLGQ-GHTPLVACDLQRPAAVYVQVGERAGVPFAPHPGASPEGPGDPVAVAA 180
QY 176 ALKAKEHLDVVIDTAGRLHIDBALMNLKVEKEIAPNEMILVDSMTFGQDAVNAE 235
DQ | : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 181 GLAEARAKHPDVVYVDTAGRLGIDIELMAQAARDAINPDEVFLVDMIGQDAVTTAA 240
QY 236 SFDDQLDVTGVTTLKLDGTGGALSTRVQPKPKFVGMSEKLDGLLELPEPMASRI 295
DQ : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 241 AFGEGVGTGVALTKUDGARGGALSVRVTGPIIFASTGEKLEDFDFVFPDMASRI 300
QY 286 LGMGDVLSLLEKAAQDDQDOEKADLEKKMRRESSFTLDDLEQLDQVKNLGLPLDIMKAP 355
DQ : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 301 LGMGDVLSLLEQAEQVFDQAQAEAAKIGAGELTLEDLFLEQLAVRKMGPIGNLLGMLP 360
QY 356 GNNKMGGLDKL-NMSEKQIDHDKALIQSMTFAENNPDLNYSRKRIAKSGSRLQEVN 414
DQ | : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 361 GAAQMK--DALAEVDDKQLDRVQALIRGMPQERADPKTIINASRLRIANGSGVTSEVN 418
QY 415 RLKQFNDMKMKMKQFTG-----GGKGGKGRNQMQNLKGNLNP 454
DQ : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
DQ 419 QLVREFFFEARKMSSMLGGMGIPGIGRKSATRKSGAKGSKGSKK-----KGTRGP 470
```

```
RESULT 8
SR54_MYCLE
ID SR54_MYCLE STANDARD; PRT; 521 AA.
AC O330J3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN FFH OR ML1622 OR MCB250.20.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE. THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z97369; CAB10614.1; -.
DR EMBL; AL583922; CAC30573.1; -.
DR HSSP; O07347; 2FFH.
DR Leproma; ML1622; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; FALSE_NEG.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 301 G-DOMAIN (BY SIMILARITY).
FT DOMAIN 302 521 M-DOMAIN (BY SIMILARITY).
FT NP_BIND 107 114 GTP (BY SIMILARITY).
FT NP_BIND 196 200 GTP (BY SIMILARITY).
FT NP_BIND 254 257 GTP (BY SIMILARITY).
SQ SEQUENCE 521 AA; 55111 MW; 37F67B8FC61BDA8B CRC64;

Query Match 47.0%; Score 1069; DB 1; Length 521;
Best Local Similarity 45.2%; Pred. No. 1.6e-48;
Matches 222; Conservative 91; Mismatches 134; Indels 44; Gaps 6;
```



```
Db 78 FQGLTGLEAANSLKGEVLTNDTAEPKDRIRALLLEADVSLPVYRFRVQSVSDQAVG 137
Qy 63 SDVMOSLTPGQGVKIVODELTKLMGENTSNMKNKPTTVMVMVGLQGAGKTTTAGKIA 122
Db 138 MGIVRGVPDQOLVKIVHDELVLKMGGEVSELQFAKSGPTVILLAGLQGVGKTTVCAKLA 197
Qy 123 LLMRKYNKKPMIYAADIYPAALINQIVGKQIDIPYVSEGDOVKPQQIVTNALHAKKE 182
Db 198 CYL-KKQSGKSMLIAGDYRPAALDQVLVGBQGVPPYTAGTDVXPADIAKQGLKEAKK 256
Qy 183 EHLDFVIIITAGRLIDHALMNEKVEYKTAEPNIMLVDSMTGQDANVAESFDOLD 242
Db 257 NNVDVIMDTAGRLQIDGMDLKVKKFLNAPTEVLLVVDAMTQGAALVYTFNVEIG 316
Qy 243 VTGVITLKIDGTGRGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMDYL 302
Db 317 ITGAILTKLDGSRGAALSVEYSGPKVLVGRGERMEDLEPTFPDMAGRILGMDYL 376
Qy 303 SLTEKAQDVDOEKADLEKAKRESSEFTLDDLEQLDQVKNLGPLDDIMKMIKPMNKKG 362
Db 377 SFVEKATEYMRQEDAEEDLQKKIMSAKPFDFNDELKQTRAVAKMGSNTRVLGNIKPGMKVSP 436
Qy 363 LDKLNMSEKQIDHIAKAIQSMTPAERNPDNLNVS--RKRRTAKSGRSLOEVNRLMKQF 420
Db 437 A-QIREAKNLLVMEAMLEVMTPEERPELLAESPEREKRIADSGKTEQVYSALVAQI 495
Qy 421 NDMKKMMKQFTGGKG 436
Db 496 FQMRVKKNLGMVMEG 511

RESULT 10
SR54_AQUAE STANDARD; PRT; 454 AA.
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN FPH OR AQ_1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN FPH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; AEO00753; AAC07579.1; -.
DR HSP; O07347; lFPH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGREMS; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 289 G-DOMAIN.
FT DOMAIN 290 454 M-DOMAIN.
FT NP_BIND 102 109 GTP (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
SQ SEQUENCE 454 AA; 50911 MW; 9f13fa8c883a0558 CRC64;

Query Match 44.4%; Score 1009; DB 1; Length 454;
Best Local Similarity 45.0%; Pred. No. 1.7e-45;
Matches 194; Conservative 103; Mismatches 8; Gaps 3;

Qy 4 EGLSERLQATMOKMGKGLTEADIKIMREVRLALFADVNFKVYKEFIKTSERALGS 63
Db 3 ELLTEKFSHALEKLTNARKITEKNINQTLREILALLEADVDYQVAKDFIKRIKRVK 62
Qy 64 DVMOSLTGPGQKIVODELTKLMGENTSNMKNKPTTVMVMVGLQGAGKTTTAGKIAL 123
Db 63 EYFKNLSPEAFVIKTVYBELVDILGGEKADLKKG-----TVLFGVLOGSKTTTIGKIAN 117
Qy 124 LMRKKYNKKPMLVAADIYRPAALINQIVGKQIDIPVYSEGDOVKPQQIVTNALKHAKKE 183
Db 118 LL-KKGHKVSVSTDLRPAAMLQQLAERLVGVYEFEBGLGAVEIARRAVKRAKEE 176
Qy 184 HLDFVIIDTAGRLHDEALMNEKEVETIAKPNELMLVDSMTGQDANVAESFDOLDV 243
Db 177 SVDYLLDITAGRLHYDELMKELEKEVETNPSEILYVADAMOGTALETAKTFPHERLGL 236
Qy 244 TGYTLKLDGDRGGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMDGVLS 303
Db 237 TGVYTKMGDARGGLALSVEVLGVPIKFIGVGEKIEDIEFFYPDRIAQRLGLGDIQS 296
Qy 304 LIEKAQDVDOEKAKDLEKMKRESSEFTLDDLEQLDQVKNLGPLDDIMKMIKPMNKKMGL 363
Db 297 LVEKAQEVITDKAQVMAKVTMTGEFDLEDLEMLRMIQMGPLDKLLSMIPGVAPO--L 354
Qy 364 DKLNMSEKQIDHIAKAIQSMTPAERNPDNLNVSRRKKRIAKSGRSLOEVNRLMKQFNDM 423
Db 355 KHLKVDKQFKKIEALINSMTPEERNRPNKTIINSRKKRIARGSGTTVSDVNKLKRYEEM 414
Qy 424 KMMKMQFTGGG 434
Db 415 KMMRKIQKAG 425

RESULT 11
SR54_BUCAI STANDARD; PRT; 451 AA.
ID SR54_BUCAI
AC P57473; Q914J1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN FPH OR BU393.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL; U82109; AAB58502.1; -.
DR PDB; 1FFH; 31-DEC-97.
DR PDB; 2FFH; 16-JUL-99.
DR PDB; 1NG1; 13-MAR-00.
DR PDB; 2NG1; 13-MAR-00.
DR PDB; 3NG1; 13-MAR-00.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54.1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54.1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 1 291 G-DOMAIN.
FT DOMAIN 292 429 M-DOMAIN.
FT NP_BIND 104 111 GTP (BY SIMILARITY).
FT NP_BIND 186 190 GTP (BY SIMILARITY).
FT NP_BIND 244 247 GTP (BY SIMILARITY).
SQ SEQUENCE 429 AA; 47225 MW; 26D387A5A303EFBD CRC64;

Query Match 44.2%; Score 1005.5; DB 1; Length 429;
Best Local Similarity 47.5%; Pred. No. 2.4e-45;
Matches 209; Conservative 92; Mismatches 118; Indels 21; Gaps 6;

QY 3 FEGLSERLQATMQKMGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERALG 62
D 1 FQQLSARLQEAIGRLGRGRITEEDLKATREIRRALMDADVNLEAVDFVVRVEEALG 60

QY 63 SDVMSQSLTPGQGVKIVQDELTKLMGENTSIINMSKPPVVMVGLQAGKTTACKLA 122
D 61 KOVLESUTPAEVLIAATVEALKEALGGEARLPVLKDR--NLWFLVGLQSGKTTAAKLA 118

QY 123 LLMEKKYKPKMLVAADIYRPAAINQLOTGVKQIDIPVSEGDQVKPQOIVTNALKHAKE 182
D 119 LYYKKG-GRPLLYAADTORPAAREQLRLGKGVGVLEVMDGESPSIRRYVEEKARL 177

QY 183 BELFDVIDTAGRLHIDEALMNEKVEKIAKPNEMLMVDSMTGODAVNVAESFDQOLD 242
D 178 EARDLILVDTAGRLQIDELPMGELARLKEVLGPDVILLDAMTQGEALSVARAFDEKVG 237

QY 243 VTGVLTKLDGTRGGAALSIRSVTKPKIFGVGSEKLDGLELHPERMA5RILGMDVL 302
D 236 VTGLVTLKLDGARGGAALSARHVTGKPIYFAGVSEKPEGLEFPFERLAGRILGMDVA 297

QY 303 SLIEKAQDVQEKAKDLEKKMRESS--FTLDDFLEQLDQVKNLGAPLDIDIMKMPGNKM 360
D 298 SLAEKV-----RAAGLAARAPKSAKELSLEDLFKQMNKLGLFPFSEILGLPGVPQ- 349

QY 361 KGLDKLMSKQIDHKAIIQSMTPAERNPDTLNVRKKRIAKSGRSLSQEVNRLMKQF 420
D 350 ---GLKVDKAIKLEAVLSMTPEERKDPRIILNGSRKKRIAKSGTSVQEVNRFKAF 405

QY 421 NDMKMMKQFTGGGKKKK 440
D 406 EEMKALMSL-----EKKKGR 421

RESULT 13
ID SR54_RICPR STANDARD; PR7; 449 AA.
AC Q9ZD20;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DE Signal recognition particle protein (Fifty-four homolog).
GN GN FFH OR RPL73.
OS OS Rickettsia prowazekii.
OC OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX OX NCBL_taxID=782;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=Madrid E;
RA RA MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria.";
RL Nature 396:133-140(1998).
CC CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC PROTEINS.
CC RUBSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

EMBL; AJ235270; CA014640.1; -.
DR HSSP; C07347; 1FFH.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54_N; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54.1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 296 G-DOMAIN.
FT DOMAIN 297 449 M-DOMAIN.
FT NP_BIND 109 116 GTP (BY SIMILARITY).
FT NP_BIND 191 195 GTP (BY SIMILARITY).
FT NP_BIND 249 252 GTP (BY SIMILARITY).
SQ SEQUENCE 449 AA; 49279 MW; 6D80817746BFA045 CRC64;

Query Match 42.2%; Score 958.5; DB 1; Length 449;
Best Local Similarity 44.2%; Pred. No. 6.8e-43;
Matches 200; Conservative 99; Mismatches 135; Indels 19; Gaps 8;

QY 3 FEGLSERLQATMQKMGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERALG 62
D 2 FXTLTQNLTKTFDKLYNSGLITENQIDTAMRDVRAVLESVALPVKGFIEVKQKALG 61

QY 63 SDVMSQSLTPGQGVKIVQDELTKLMG--ENTSIINMSKPPVVMVGLQAGKTTAGK 120
D 62 QEVIKSVSPGMIKTIHEEMINLLASTESTKLNNAKPPVNLIVLQGGGKTTASGK 121

QY 121 LALLMRKYNKKMPLVAADIYRPAAINQLOTGVKQIDI---PVYSEGDVQKQIVNAL 177
D 122 LALLR-KNQNKVLLVSLDTPYRAAQEQELAIANSVINDSLPIV-KGE--KPLDIVKRAI 177

QY 178 KHAKHEHLDFVIDTAGRLHIDEALMNEKVEKIAKPNEMLMVDSMTGODAVNVAESF 237
D 178 GEAQISADVVIDTAGRTQIDNVNMEALAIKILNPTETLLVIDSMTCQDAVITAKTF 237

```



```
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39684; AAC71264.1; -.
DR HSP; O07347; 1FTH.
DR TIGR; MG048; -.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPP; 1.
DR ProDom; PD000819; SRP54; 1.
DR TrRFams; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 293 G-DOMAIN.
FT DOMAIN 294 446 M-DOMAIN.
FT NP_BIND 106 113 GTP (BY SIMILARITY).
FT NP_BIND 188 192 GTP (BY SIMILARITY).
FT NP_BIND 246 249 GTP (BY SIMILARITY).
SQ SEQUENCE 446 AA; 50199 MW; 080F225C8B5E5AE CRC64;

Query Match 39.5%; Score 898; DB 1; Length 446;
Best Local Similarity 43.0%; Pred. No. 8.8e-40;
Matches 198; Conservative 90; Mismatches 135; Indels 38; Gaps 10;

QY 6 LSERLQATQMKRKGKGLTEADIKIMREYRLALFEADYVKNVKEFIKTVSERALGSDY 65
DB 6 LSSIVMTQKKNQATTEKDVLEVLKELRIALLDADNLLVKNFIKAIKDTKVG--- 62

QY 66 MQSLTPGQ---QVIKIVQDELTKMGENTSINMSNKPPTVVMVGLQGAGKTTAGKL 121
DB 63 -QTIEPGQDLQKSLKTIKTELINLSQPQELN--EKRLKIMVGLQSGSKTTGCKL 119

QY 122 ALLMRKYNKPMVAADIIYPAAINOLQTVGKQIDIPVYSEGQ--VKPQIIVTNALXH 179
DB 120 AYWLKRYKQKTMVLGDIYRPAALQETLSQQTNSVFFAQGTQFVAKTTKAALSFAKT 179

QY 180 AKESHLDFVIIDFAGRLHIDALMNLKVEKTAKEINIMLVDSMTGODAVNVASFD 239
DB 180 AK--COTICDTAGRLQTNELMDELVSVKNEINPEIIMVVDLSGQELINVAQTFHK 236

QY 240 QLDVYGLTKLDGTRGSAALSIRSVTQPKIFVGMSEKLDGLDFHPERMASRILGMG 299
DB 237 RLKITGFIISKLSDARAGAALSASLQVPKILGVSEKLDGLEQHPERINRIILGLG 296

QY 300 DVLSLEPKAQDVDEKANDLEK--KWRSSSTLDDFLEQLDVKNLGLPDDIMKNMIPG 356
DB 297 DVMSLVEKAQVFDK---KDLTKTISKMFLGKMDLEDLLIYMQQMHKMGVSSLIKMLPA 353

QY 357 MNKMKGLDKLNSEKQIDHIKAIQSMTPAERNPDTLN--VSRKRIKAGSGHSLQEVN 414
DB 354 -NFSVSENELIENKIELWKLVTNSITREEREPKLIINRDPNRRKRIKSGSKMDL 412

QY 415 RLMQFNDMKMKMKQFTGGGKKGKKNQNMKGNLPLF 455
DB 413 KLMKKNKQL-----KATEMKLLKTSNPF 439
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:15:12 : Search time 35 Seconds
(without alignments)
2678.612 Million cell updates/sec

Title: US-09-943-108a-2
Perfect score: 2274
Sequence: 1 MAFGLSERLQATMQRGK.....GKKGRNQNMILKGNLPLF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_plant.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	2274	100.0	455	16	Q99UN3	Q99un3 staphylococ
2	1717.5	75.5	450	16	Q8Y695	Q8y695 listeria mo
3	1708.5	75.1	450	16	Q92AK7	Q92ak7 listeria in
4	1584	69.7	451	16	Q9KA10	Q9ka10 bacillus ha
5	1393	61.3	449	16	Q97L98	Q97l98 clostridium
6	1382	60.8	452	16	Q8XJP3	Q8xjp3 clostridium
7	1352.5	59.5	447	16	Q8R9X0	Q8r9x0 thermoanaer
8	1315	57.8	518	16	Q9CF65	Q9cf65 lactococcus
9	1306	57.4	523	16	Q97GD2	Q97gd2 streptococc
10	1279.5	56.3	519	16	Q982K1	Q98zk1 streptococc
11	1238	54.4	444	16	Q8RDV7	Q8rdv7 fusobacteri
12	1180.5	51.9	433	16	Q9XIQ1	Q9xiq1 thermotoga
13	1173.5	51.6	453	16	Q8XF48	Q8xf48 salmonella
14	1150.5	50.6	461	16	Q9KUG1	Q9kug1 vibrio chol
15	1150	50.6	457	16	Q9HXP8	Q9hxp8 pseudomonas
16	1139.5	50.1	453	16	Q8ZBU6	Q8zbu6 yersinia pe

17	1135.5	49.9	458	16	Q9CLN6	Q9CLN6 pasteurella
18	1131	49.7	456	16	Q9K177	Q9K177 neisseria m
19	1130	49.7	456	2	Q9JPK6	Q9JPK6 neisseria m
20	1130	49.7	456	16	Q9JSQ0	Q9JSQ0 neisseria m
21	1123	49.4	456	2	Q8RNK2	Q8RNK2 neisseria g
22	1121.5	49.3	458	2	Q9L9C9	Q9L9C9 thiobacillu
23	1092.5	48.4	476	16	Q8XVL9	Q8XVL9 ralstonia s
24	1092.5	48.0	490	16	Q8VVM3	Q8VVM3 anabaena sp
25	1081.5	47.6	532	16	Q98E72	Q98E72 rhizobium i
26	1077.5	47.4	550	16	Q69874	Q69874 streptomyce
27	1070.5	47.1	550	2	O85676	O85676 streptomyce
28	1066.5	46.9	452	16	Q9RTC9	Q9RTC9 deinococcus
29	1062.5	46.7	554	16	Q8UC01	Q8UC01 agrobacteri
30	1054	46.4	523	16	Q8YJ61	Q8YJ61 brucella me
31	1050.5	46.2	449	16	Q98QL9	Q98QL9 mycoplasma
32	1050.5	46.2	513	16	Q92L45	Q92L45 rhizobium m
33	1039	45.7	513	2	Q8VVP8	Q8VVP8 arcanobacte
34	1018	44.8	434	16	Q9DPH7	Q9DPH7 xylella fas
35	981.5	43.2	449	16	Q92J55	Q92J55 rickettsia
36	981	43.1	525	10	O82532	O82532 pisum sativ
37	970.5	42.7	508	16	Q9A2B3	Q9A2B3 caulobacter
38	933.5	41.1	448	16	Q9PR53	Q9PR53 ureaplasma
39	905.5	39.8	447	16	O51637	O51637 borrelia bu
40	899	39.5	549	10	Q9AXU1	Q9AXU1 chlamydomon
41	888	39.1	448	16	Q9PL14	Q9PL14 chlamydia m
42	869	38.2	448	16	O84028	O84028 chlamydia t
43	846	37.2	448	16	Q9JRV9	Q9JRV9 chlamydia p
44	846	37.2	448	16	Q9Z966	Q9Z966 chlamydia p
45	843	37.1	445	16	Q9PPJ8	Q9PPJ8 campylobact

ALIGNMENTS

RESULT 1

Q99UN3 PRELIMINARY; PRT; 455 AA.

AC Q99UN3

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Ffh protein (Signal recognition particle homolog).

GN FFH OR SAV1237 OR SAL080.

OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and

OS Staphylococcus aureus (strain N315).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;

OC Staphylococcus.

OX NCBI_TaxID=158878, 158879;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);

RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,

RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,

RA Mizutani Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,

RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,

RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;

RT "Whole genome sequencing of methicillin-resistant Staphylococcus aureus.";

RL Lancet 357:1225-1240(2001).

DR EMBL; AP003361; BAB57399.1; -.

DR EMBL; AP003133; BAB42332.1; -.

DR HSSP; O07347; 1FFH.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR000897; SRP54.

DR InterPro; IPR004125; SRP54_SBP.

DR InterPro; IPR004780; SRP_sub.

DR Pfam; PF00448; SRP54; 1.

DR Pfam; PF02881; SRP54.N; 1.

DR Pfam; PF02978; SRP_SBP; 1.

DR ProDom; PD000819; SRP54; 1.

```
DR SMART; SW00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 455 AA; 50753 MW; 619A202B1734097C CRC64;

Query Match
Best Local Similarity 100.0%; Score 2274; DB 16; Length 455;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAPEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
QY 121 LALLMRKYNKPKMLVAADIYRPAALINQLTQVKGIDIPYVSEGDQVKKPQQIVTNALKHA 180
DB 121 LALLMRKYNKPKMLVAADIYRPAALINQLTQVKGIDIPYVSEGDQVKKPQQIVTNALKHA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDQ 240
DB 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDQ 240
QY 241 LDVGTGVTIKLGDTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFPHPERMASRILGMGD 300
DB 241 LDVGTGVTIKLGDTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFPHPERMASRILGMGD 300
QY 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
DB 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
QY 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLSQEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLSQEVNRLMKQF 420
QY 421 NDMKKMKQFTGGGKGKGRNQMNKLGKMLPF 455
DB 421 NDMKKMKQFTGGGKGKGRNQMNKLGKMLPF 455

RESULT 2
QY695 PRELIMINARY; PRT; 450 AA.
AC QY695;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Ffh protein.
GN FFH OR LMO1801.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
[1]
SEQUENCE FROM N.A.
STRAIN=EGD-E / SEROVAR 1/2A;
MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fiehl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maifournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluter T., Simoes N., Tisseret A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
SCIENCE 294:849-852(2001).
```

```
DR EMBL; AL591981; CAC99879.1; -.
DR ListList; LMO01801; -.
DR InterPro; IPR003593; AAA_AtpPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 49714 MW; D5888D7A99C60C76 CRC64;

Query Match
Best Local Similarity 75.5%; Score 1717.5; DB 16; Length 450;
Matches 331; Conservative 64; Mismatches 55; Indels 5; Gaps 1;

QY 1 MAPEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAPEGLSERLQATMOKMRGKGLTEADIKIMREVRLLALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
DB 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSNMKNPPTVVMVGLQGAGKTTTGGK 120
QY 121 LALLMRKYNKPKMLVAADIYRPAALINQLTQVKGIDIPYVSEGDQVKKPQQIVTNALKHA 180
DB 121 LALLMRKYNKPKMLVAADIYRPAALINQLTQVKGIDIPYVSEGDQVKKPQQIVTNALKHA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDQ 240
DB 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKETAKPNEIMLVDSMTGQDAVNVASFDQ 240
QY 241 LDVGTGVTIKLGDTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFPHPERMASRILGMGD 300
DB 241 LDVGTGVTIKLGDTRGGAALSIRSVTKPIKFGVMSSEKLDGLFELFPHPERMASRILGMGD 300
QY 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
DB 301 VLSLIEKAQVDQEKAKOLEKKMRSEFTLDDFLEQLDQVKNLGPLDDIMKPIGMNKM 360
QY 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLSQEVNRLMKQF 420
DB 361 KGLDLNMSKQIDHKAIIQSTPAERNPDTLNVSRKRIAKGSSRSLSQEVNRLMKQF 420
QY 421 NDMKKMKQFTGGGKGKGRNQMNKLGKMLPF 455
DB 421 NDMKKMKQFTGGGKGKGRNQMNKLGKMLPF 455

RESULT 3
QY2AK7 PRELIMINARY; PRT; 450 AA.
AC QY2AK7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ffh protein.
GN FFH OR LIN1915.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
[1]
SEQUENCE FROM N.A.
STRAIN=CLIP 11262 / SEROVAR 6A;
RC PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
```

```

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Raestri U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97145.1; -.
DR L101181; L101181; -.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 49684 MW; E417626DID69D239 CRC64;

Query Match 75.1%; Score 1708.5; DB 16; Length 450;
Best Local Similarity 72.3%; Pred. No. 9.7e-84;
Matches 329; Conservative 65; Mismatches 56; Indels 5; Gaps 1;

QY 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSVDVMSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQGAGKTTTAK 120
DB 61 VGDVMSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQGAGKTTTAK 120
QY 121 LALLMRKYNKPMVAADYIPRAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADYIPRAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
QY 181 KEEHLDVVIDTAGRLHIDEALMELKEVEIAKPNEMLVVDSMTGQDVAVNAESFDDQ 240
DB 181 KEEHLDVVIDTAGRLHIDEALMELKEVEIAKPNEMLVVDSMTGQDVAVNAESFDDQ 240
QY 241 LDVTGVTLTKLGDTRGGAALSIRSVTKPIKFGVSEKLDGLFELFHPERMASRIILMGD 300
DB 241 LEITGVTLTKLGDTRGGAALSIRSVTKPIKFGVSEKLDGLFELFHPERMASRIILMGD 300
QY 301 VLSLIEKAQDVQDEKADLEKKMRESFTLDDFLQDLQVKNLGPLDDIMKMPGNMKM 360
DB 301 VLSLIEKAQDVQDEKADLEKKMRESFTLDDFLQDLQVKNLGPLDDIMKMPGNMKM 360
QY 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKRRIAKSGRSLOEVNRLMKQF 420
DB 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKRRIAKSGRSLOEVNRLMKQF 420
QY 421 NDMMKKMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455
DB 421 AEMKKMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455

RESULT 4
Q9KAL0 PRELIMINARY; PRT; 451 AA.
AC Q9KAL0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Signal recognition particle.
GN FFH OR BH2484.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001515; BAB06203.1; -.
DR HSSP; O07347; 1FFH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 451 AA; 50104 MW; 1B61BEP45E14674F CRC64;

Query Match 69.7%; Score 1584; DB 16; Length 451;
Best Local Similarity 67.3%; Pred. No. 4.4e-77;
Matches 307; Conservative 73; Mismatches 70; Indels 6; Gaps 2;

QY 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSVDVMSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQGAGKTTTAK 120
DB 61 LGQVMSLTGPGQVQIKIVODELTKLMGENTSIKNSKPPPTVVMVGLQGAGKTTTAK 120
QY 121 LALLMRKYNKPMVAADYIPRAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
DB 121 LALLMRKYNKPMVAADYIPRAINQLQTVGKQIDIPVYSEGQVQKQOIVTNALKHA 180
QY 181 KEEHLDVVIDTAGRLHIDEALMELKEVEIAKPNEMLVVDSMTGQDVAVNAESFDDQ 240
DB 181 KEEHLDVVIDTAGRLHIDEALMELKEVEIAKPNEMLVVDSMTGQDVAVNAESFDDQ 240
QY 241 LDVTGVTLTKLGDTRGGAALSIRSVTKPIKFGVSEKLDGLFELFHPERMASRIILMGD 300
DB 241 LDVSGVTLTKLGDTRGGAALSIRSVTKPIKFGVSEKLDGLFELFHPERMASRIILMGD 300
QY 301 VLSLIEKAQDVQDEKADLEKKMRESFTLDDFLQDLQVKNLGPLDDIMKMPGNMKM 360
DB 301 VLSLIEKAQDVQDEKADLEKKMRESFTLDDFLQDLQVKNLGPLDDIMKMPGNMKM 360
QY 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKRRIAKSGRSLOEVNRLMKQF 420
DB 361 KGLDKLNMSEKQIDHKAIIQSMTPAERNPDPLNYSRKRRIAKSGRSLOEVNRLMKQF 420
QY 421 NDMMKKMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455
DB 421 NDMMKKMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455
DB 421 EEMKKMKQFTG-GGKGGKGRNQNMNMLKGMNLPF 455

RESULT 5
Q97198 PRELIMINARY; PRT; 449 AA.
AC Q97198;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Signal recognition particle.
GN CAC1754.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Clostridia;
RN [1]

```

```
OC Clostridiales, Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007684; AAK79720.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 449 AA; 49555 MW; FF4695E75A673733 CRC64;

Query Match 61.3%; Score 1393; DB 16; Length 449;
Best Local Similarity 61.0%; Pred. No. 6.9e-67;
Matches 280; Conservative 71; Mismatches 94; Indels 14; Gaps 5;

QY 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREYRLALFEADVNFVKYKEFIKTVSRA 60
DB 1 MAFEGSLTQLQAAMKLRGKGLSEKIDKAMREYKLLALFEADVNFVKYKEFIKTVSRA 60

QY 61 LGSVDVMSLTGQGVKIVQDELTKLGGENTISNNKPTVVMVGVGGAGKTTAGK 120
DB 61 LGNEVRESLTGQGVKIVNEELTNLGGESKIEFEENGITVIMAVGLGGAGKTTAGK 120

QY 121 LALLMRKYNKPMVAADYRPAALNQLQTVGKQIDIPVYSEGQDVKKQCIIVTNALKHA 180
DB 121 LSLSLKKK-NKKPLLVACDIYRPAALKQLEVYKSIDVPVSMGDKVNPVDSKAAMKHA 179

QY 181 KEEHLDVFIIDTAGRLHIDBALMNEKVEYKTAKEPNEIMLVDSMTGQDAVNVASFDDQ 240
DB 180 KENGLNVYIIDTAGRLHIDBALMNELENTKSEVNPKEILLVYDSMTGQDAVNVASFDDK 239

QY 241 LDVTGVTITKLDGTRGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMGD 300
DB 240 LELTGVLTITKLDGTRGAALSIREMTGPIKYGLGKMDNIEIFHPDEMASRILGMGD 299

QY 301 VLSLIEKAQDVQDEKAKLEKMRSSFTLDDFLEQLDQVKNLGLDDIMKMPGMNK- 359
DB 300 VLTLLIEKAQAIDEKQAKELGDRMLSQEFNFDFLQAPEQMKLGPGLKLEWVPGFNSS 359

QY 360 -MKGDLKNMSEKQIDHIKAIQSMTPAERNPD--TLNYSKRRTAKSGRSRLOEVNRL 416
DB 360 MLKGYD-LSKNEGEMKYRIIISMTAKERKNPSLITSTASRRRIALSGTGVQEVNRI 418

QY 417 MKQFNDMKMKMQFTGGGKGGKRNQOMNMLKGNLNF 455
DB 419 LKNEFQMKMKMQFTGGKGNKFSKG-----LFGGKMFF 449

RESULT 6
Q8XJP3 PRELIMINARY; PRT; 452 AA.
AC Q8XJP3;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Signal recognition particle protein.
GN PFH OR CPEL713.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ontani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003191; BAB81419.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 452 AA; 49685 MW; 2356112EDC9C7D31 CRC64;

Query Match 60.8%; Score 1382; DB 16; Length 452;
Best Local Similarity 62.7%; Pred. No. 2.7e-66;
Matches 279; Conservative 72; Mismatches 86; Indels 8; Gaps 5;

QY 1 MAFEGSLERLQATMOKRGKGLTEADIKIMREYRLALFEADVNFVKYKEFIKTVSRA 60
DB 1 MAFDGLASKLQDTLKLKGKGLTEKIDKEAMREYKLLALFEADVNFVKYKEFIKTVSRA 60

QY 61 LGSVDVMSLTGQGVKIVQDELTKLGGENTISNNKPTVVMVGVGGAGKTTAGK 120
DB 61 VGEVLNLSLTGQGVKIVNDELTTLMGETESKLYSDNGPTVFMVLVGLGGAGKTTAGK 120

QY 121 LALLMRKYNKPMVAADYRPAALNQLQTVGKQIDIPVYSEGQDVKKQCIIVTNALKHA 180
DB 121 LALLMRK-NKKPLLVACDIYRPAALKQLEVYKSIDVPVSMGDKVNPVDSKAAMKHA 179

QY 181 KEEHLDVFIIDTAGRLHIDBALMNEKVEYKTAKEPNEIMLVDSMTGQDAVNVASFDDQ 240
DB 180 KDNNGNVYIIDTAGRLHIDBALMNELENTKSEVNPSEILLVVDAMTQDAVNVASFENNS 239

QY 241 LDVTGVTITKLDGTRGAALSIRSVTKPIKFGVMSKLDGLLELPHPERMASRILGMGD 300
DB 240 LELSGIILITKLDGTRGAALSIRDTGKPIKFGVMSKLDGLLELPHPERMASRILGMGD 299

QY 301 VLSLIEKAQDVQDEKAKLEKMRSSFTLDDFLEQLDQVKNLGLDDIMKMPGMNK-- 358
DB 300 VLSLIEKAQAIDDEASKLSERMLNQEFNFDYLSAMDQMKLGPINKLIEMPGVNTK 359

QY 359 MKGDLKNMSEKQIDHIKAIQSMTPAERNPD--NLVSRKRRTAKSGRSRLOEVNRL 416
DB 360 ELEGID-FSQGEKQMATVKAIQSMTAKERKQPSLVIGNSKRRTAKSGTGVQEVNRY 418

QY 417 MKQFNDMKMKMQFTGGGK--GKGG 439
DB 419 LKGYEMMKMKMQFTGGGK 443

RESULT 7
Q8RX90 PRELIMINARY; PRT; 447 AA.
ID Q8RX90
AC Q8RX90;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
```

01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
Signal recognition particle GTPase.
FFH OR TrEMBL62.
Thermoanaerobacter tengcongensis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
NCBI_TaxID=119072;
[1]
SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013104; AAM24584.1; -.
KW Complete proteome.
SQ SEQUENCE 447 AA; 49294 MW; 3DB00C468FA14B7E CRC64;

Query Match 59.5%; Score 1352.5; DB 16; Length 447;
Best Local Similarity 59.9%; Pred. No. 1e-64;
Matches 273; Conservative 74; Mismatches 96; Indels 13; Gaps 4;

QY 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRLALFEADVNFVKVKEFIKTVSERA 60
DB 2 MAFESLSERLQGVFKLRGKGLTEKDIKEMREVKVALLADVNFVKVGFINSVTEKA 61
QY 61 LGSVMOSLTPGGQVQIKIVQDELTKLGGENTSIINMSNKPPTVVMVGLQGAGKTTAGK 120
DB 62 LQGEVMSLTPAQOVQIKIVHEELIKLGSVESRLNLGSKVPVIMVGLQSGKTTACGK 121
QY 121 LALLMRKYNKPKMLVAADIYPPAIIQVGVQIDIPVYSEGQVQPKQIVTNALKHA 180
DB 122 LANLL-KQGNPLLVACDTVPAAIKQLVLGANINVPFTMGDKVDTADIAKASDYA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNELEKVEIAKPNELMIVDSMTGQDVANVAESFDDQ 240
DB 181 KSHNVVDVIIDTAGRLHIDEALMNELEKVEIAKPNELMIVDSMTGQDVANVAESFDDQ 240
QY 241 LQVGVTLTKLDGDTGGGAALSIRSVTKPIKFKVGMSEKLDGLFELHPERMASRILGMGD 300
DB 241 LDITGVTLTKLDGDTGGGAALSIRSVTKPIKFKVGMSEKLDGLFELHPERMASRILGMGD 300
QY 301 VLSLIEKAQDVQDEKAKOLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMK 360
DB 301 VLTIEKAQAAIDDKALEMGKILSKOFTLEDFLEQLSLKNWGLPOLLAMIPGVNKK- 359
QY 361 KGLDKLAMSKEQIDHIKAIQSMTPAERNPDITLVNKKRIAKSGRSLOEVNRLMKQF 420
DB 360 SVLNVSEKDLKRIEAILLSTKEEQNPSTINGSKRRIARGSGTIOEVNRLMKQF 419
QY 421 NDMKMMKQFTGGGKG-KXGKRNMQNMLKGNLPP 455
DB 420 EETKMMKRFADIKDLKRGK-----LRLPF 445

RESULT 8
ID Q9CF65 PRELIMINARY; PRT; 518 AA.
AC Q9CF65;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Signal recognition particle protein Ffh.
GN FFH OR L1616.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jailion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006391; AK05714.1; -.
DR HSP; A007347; IFFH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 518 AA; 56845 MW; BF4539EF037601B4 CRC64;

Query Match 57.8%; Score 1315; DB 16; Length 518;
Best Local Similarity 55.4%; Pred. No. 1.2e-62;
Matches 258; Conservative 88; Mismatches 102; Indels 18; Gaps 2;

QY 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRLALFEADVNFVKVKEFIKTVSERA 60
DB 1 MAFENLTERLQNVFKLRGKKTETDVTETKTSIRVALLEADVLPVVKFIKAIERA 60
QY 61 LGSVMOSLTPGGQVQIKIVQDELTKLGGENTSIINMSNKPPTVVMVGLQGAGKTTAGK 120
DB 61 VGEVSESLNPAQOVQIKIVDEELTAILGGGEAEELKSPKPTIIMVGLQAGKTTAGK 120
QY 121 LALLMRKYNKPKMLVAADIYPPAIIQVGVQIDIPVYSEGQVQPKQIVTNALKHA 180
DB 121 LAKLKEQNARPLMTAADVYPPAIDQLTKLGSLPVDYDGTAEKPVNIVRNLKA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNELEKVEIAKPNELMIVDSMTGQDVANVAESFDDQ 240
DB 181 QBERKDYVLIDTAGRLHIDEALMNELEKVEIAKPNELMIVDSMTGQDVANVAESFDDQ 240
QY 241 LQVGVTLTKLDGDTGGGAALSIRSVTKPIKFKVGMSEKLDGLFELHPERMASRILGMGD 300
DB 241 LDITGVTLTKLDGDTGGGAALSIRSVTKPIKFKVGMSEKLDGLFELHPERMASRILGMGD 300
QY 301 VLSLIEKAQDVQDEKAKOLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMPGNMK 360
DB 301 VLTIEKAQAAIDDKALEMGKILSKOFTLEDFLEQLSLKNWGLPOLLAMIPGVNKK- 359
QY 361 KGLDKLAMSKEQIDHIKAIQSMTPAERNPDITLVNKKRIAKSGRSLOEVNRLMKQF 420
DB 361 PGLDKVKVDFKDVARKAMVLSNTPAERHLEAELSPARRRRIAGSGNSFIEVNFQF 420
QY 421 NDMKMMKQFTGGGKG-KXGKRNMQNMLKGNLPP 454
DB 421 NOSKEMMQGIMNGDNAMQKMMGGAGG-----ONPNMPAGSGMP 460

RESULT 9
ID Q97QD2 PRELIMINARY; PRT; 523 AA.
AC Q97QD2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Signal recognition particle protein.
GN SPI287.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angicoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007428; AAK75391.1; -.
DR TIGR; SP12871; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 523 AA; 57757 MW; A31745B7898470F2 CRC64;

Query Match 57.4%; Score 1306; DB 16; Length 523;
Best Local Similarity 54.9%; Pred. No. 3.7e-62;
Matches 252; Conservative 91; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSRLQATMOKMRGKGLTEADIKIMREVRALFEADVNFVKYKEIKTVSSRA 60
Db |||||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MAFESLTERLQWFKNLKRGKTSDESVDQEAATKEIRLALLAEADVLPVVKDIKKVRRA 60
QY 61 LGSVMQSILTPGQOVIRIVODELTKLGGENTSIKNSKPTVYVMVGLQGAGKTTAGK 120
Db |||||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VGEVIDTLNPAQQLIKIVDEETAVLGSDTAETIKSPKPTIIMVMVGLQGAGKTTAGK 120
QY 121 LALLMRKYNKKPMLVAADIYRPAALNQLQTVGKQIDIPYSEGDQVKPQOIVTNALXHA 180
Db ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 121 LANKLKEENARPLMIAADIYRPAADQLKLGQIDVPVFGALGTEVPAVEIVRGLSEA 180
QY 181 KEEHLDVLIIDTAGRLHIDEALMNEKVEKTAEPNEIMLVDSMTGQADVNVASFDDQ 240
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 181 QTNHNDYVLIDTAGRLQIDELLMNEURDYKALAQPNETILLVDAMTIGOEAAVREFNAQ 240
QY 241 LDVTGVTLTLDGTRGGAALSIRSVTQPKIFVGMSEKLDGLFHPERMASRIILMGD 300
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 241 LEVTVGLTKLDGTRGGAALSIRSVTQPKIFGTGTEKIDTETHPDMSRILMGD 300
QY 301 VLSLIEKAQDVQDEKAKLEKXMRSSFTLDDLEQDQVKNLGLPDDIMKMFPMNMK 360
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 301 MLTLIEKASQETDEKKALEAAEKMRNTDFDNDQDQVQNGMPWDLKXIPGMANN 360
QY 361 KGLDKLNKSEKQIDHIKAIQTOSTPAERNPDTLNVSKKRIAGSGRSLOEVRNLKQF 420
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 361 PALQNMKVDKQIARKRAIVSSMTPAERENPDLLNPSRRRIAGSGNTFVEYNKFIKD 420
QY 421 NDMMKMKQFTGGGKGKGRNQNMLKGM-----NLP 454
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 421 NQAKQLMQGVMSG-----DNKMKMQMGINPNLP 450

RESULT 10
Q99ZK1 PRELIMINARY; PRT; 519 AA.
AC Q99ZK1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
```

```

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DR Putative signal recognition particle.
GN RFH OR SP1200.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006560; AAK34060.1; -.
DR HSSP; O07347; 2FFH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 519 AA; 57415 MW; EF1E46696A1C0040 CRC64;

Query Match 56.3%; Score 1279.5; DB 16; Length 519;
Best Local Similarity 54.9%; Pred. No. 9.6e-61;
Matches 251; Conservative 86; Mismatches 113; Indels 7; Gaps 2;

QY 1 MAFEGLSRLQATMOKMRGKGLTEADIKIMREVRALFEADVNFVKYKEIKTVSSRA 60
Db |||||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MAFESLTERLQWFKHGRGKLSDESVDQEVTKERLALLAEADVLPVVKTFIKVRERA 60
QY 61 LGSVMQSILTPGQOVIRIVODELTKLGGENTSIKNSKPTVYVMVGLQGAGKTTAGK 120
Db ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 61 IGHEIIDLTPQQLIKIVNEELTSLGSETAIDKSPKPTIIMVMVGLQGAGKTTAGK 120
QY 121 LALLMRKYNKKPMLVAADIYRPAALNQLQTVGKQIDIPYSEGDQVKPQOIVTNALXHA 180
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 121 LANKLKEENARPLMIAADIYRPAADQLKLGQOINVPVFDMGTDHSAVDIYRKLBOA 180
QY 181 KEEHLDVLIIDTAGRLHIDEALMNEKVEKTAEPNEIMLVDSMTGQADVNVASFDDQ 240
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 181 RENHNDYVLIDTAGRLQIDELKNGELRDYKALAQPNETILLVDSMTIGOEAAVVEFNHQ 240
QY 241 LDVTGVTLTLDGTRGGAALSIRSVTQPKIFVGMSEKLDGLFHPERMASRIILMGD 300
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 241 LSTTGVTLTLDGTRGGAALSIRSVTQPKIFGTGTEKIDTETHPDMSRILMGD 300
QY 301 VLSLIEKAQDVQDEKAKLEKXMRSSFTLDDLEQDQVKNLGLPDDIMKMFPMNMK 360
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 301 LUTLIEKASQETDEKKALEAAEKMRNTDFDNDQDQVQNGMPWDLKXIPGMANN 360
QY 361 KGLDKLNKSEKQIDHIKAIQTOSTPAERNPDTLNVSKKRIAGSGRSLOEVRNLKQF 420
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 361 PALANIKVDENQIARKRAIVSSMTPAERENPDLLNPSRRRIAGSGNSFVD-NKFIKDF 419
QY 421 NDMMKMKQFTGGGKGKKGK-----RNQMNMLKGM 451
Db |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 420 NQAKSMQGVMSGDKMKMKMDMGINPNLPKNAPAGM 456

RESULT 11
Q8RDV7
```

```
ID Q8RDV7 PRELIMINARY; PRT; 444 AA.
AC Q8RDV7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Signal recognition particle, subunit FFH/SRP54.
GN FN1393.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21866394; PubMed=11869109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010644; AAL95586.1; -.
KW Complete proteome.
SQ SEQUENCE 444 AA; 48966 MW; F2593E0D3F0F851A CRC64;

Query Match 54.48; Score 1238; DB 16; Length 444;
Best Local Similarity 54.28; Pred. No. 1.3e-58;
Matches 245; Conservative 86; Mismatches 111; Indels 10; Gaps 3;

QY 4 EGSERLQATMQKRGKGLTEADIKIMREVRALFEADVNFVKVEFTKTVSERALGS 63
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
3 ENLGNRPDIPFKIRGHKGLSETNIKDALREVKMSLEADVNVKVDFTNIRSEKAI 62
QY 64 DVMSQSTPGQQVIKIVDELTKLMGGENTSINNSKNPPTVVMVYGLQGAGKTTT 123
Db : : : : | | | | | : | | | | | : | | | | | : | | | | | :
63 EVIRGVNPAQQTILVNDELVELLGGTSSKLTGLRNPITIMLAGLOGAGKTTFAAK 122
QY 124 LMRKKNKPMPLVAADYRPAAINQLQTVGKQIDIPVYSEGQVYKPOQIVTNAL 183
Db : : : : | | | | | : | | | | | : | | | | | : | | | | | :
123 FL-KKQEKLLIVGVYRPAAIKQLVGLQGGVGVYSEEDNKDVVGVIATRAEKAK 181
QY 184 HLDVFTIDTAGRLHIDEALMNEKVEKIAKPNEIMLVVDSMTGODAVNVAESFD 243
Db : : : : | | | | | : | | | | | : | | | | | : | | | | | :
182 NATYMLVDYAGRLHVDFTLDELKELKA-KPQEIILVVDAMIGODAVNVAESFN 241
QY 244 TGVTLFKDGDTRGGAALSIRSVTKPIKFGVMSEKLDGLFHPERMASRILGMGD 303
Db || || || || || || || || || || || || || || || || || || || ||
242 DGVILTKDGDTRGGAALSIRSVTKPIKFGVGEKINDIEFHPDRILVSRILGMGD 301
QY 304 LIEKAQDVQDEKAKDLEKKNRESFTLDDFLEQLDQVKNLGLPDDIMKMPGN 363
Db ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
302 LVKAQEVIDEENAKSLEKTSKQFQNDPLFKQLQTKRGLSGGLIKLIPGNPK- 358
QY 364 DKLMSSEKIDHIKAIQSMTPAERNPDTLVNSRKKRIAGSGRSLOEYNRLMK 423
Db | | : : : : | | | | | : | | | | | : | | | | | : | | | | | :
359 DOLAPAEKEMKVEALIQSMFKERKPKDILKASRIKIRIAGSGDVSNDYKNLL 418
QY 424 KMMKQFTGGGKGGKGRKNQNMKGNLPPF 455
Db || | | : : | | | | : : | | | | : : | | | | : : | | | | :
419 KSMKMF-----SKMPNIGAMGKGRFPF 444

RESULT 12
Q9X1Q1 PRELIMINARY; PRT; 433 AA.
AC Q9X1Q1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Signal recognition particle protein.
GN TM1565.
OS Thermotoga maritima.
```

```
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Winn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Maiek J.A., Linner K.D., Garrett M.K.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima";
RL Nature 399:323-329(1999).
DR EMBL; AE001802; AAD36632.1; -.
DR HSSP; O07347; IFFH.
DR TIGR; TM1565; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SBP.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMS; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 433 AA; 48701 MW; B907E982F4D29357 CRC64;

Query Match 51.9%; Score 1180.5; DB 16; Length 433;
Best Local Similarity 53.4%; Pred. No. 1.5e-55;
Matches 234; Conservative 92; Mismatches 99; Indels 13; Gaps 5;

QY 3 FEGSERLQATMQKRGKGLTEADIKIMREVRALFEADVNFVKVEFTKTVSERALG 62
Db || | | | : : | | | | | : | | | | | : | | | | | : | | | | | :
2 FENQEKLSRVFKSGRGKITEKNVKAIDAREVLSLEADVNVKVEFVDHVLQALG 61
QY 63 SDVMQSTPGQQVIKIVODELTCLMGGENTSINNSKNPPTVVMVYGLQGAGKTTT 122
Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
62 EEVLSLTDPQQIKIVRDELVRIMGEKNEPLVHR-PAPIMVVGLOSSGKTTTCAK 120
QY 123 LLMRKKNKPMPLVAADYRPAAINQLQTVGKQIDIPVYSEGQVYKPOQIVTNAL 182
Db ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
121 KLL-KKEGRNPLVAADLYRPAADVQLKLGNIQVNVVHDYNT-PVEIVKEADV 178
QY 183 EHLDFVITDAGRLHIDEALMNEKVEKIAKPNEIMLVVDSMTGODAVNVAES 242
Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
179 TGDVLIVDTAGRLHIDEEMKELEEKILNDEILLVVDAMGQDVAVTAKVDF 238
QY 243 VTGVTFLKDGDPTRGGAALSIRSVTKPIKFGVMSEKLDGLFHPERMASRIL 302
Db : : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
239 LTGSPVTKMDGARGVILSIKIVTGKPVKFIGTSEKLDGLEPFPDRIANRIL 298
QY 303 SLIEKAQDVQDEKAKDLEKKNRESFTLDDFLEQLDQVKNLGLPDDIMKMP 362
Db || || | : : | | | | | : | | | | | : | | | | | : | | | | | :
299 SLIEKVEKLDQDEKMKSAEKFLAAEFTLFDFREQLQEMKGLGPLSSILE 356
QY 363 LDKLMSSEKIDHIKAIQSMTPAERNPDTLVNSRKKRIAGSGRSLOEYNRL 422
Db : | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
357 --DVMSSEKELKIEALINSMITIERRNPGLIINASRKRIRIAGSGTTVD 414

QY 423 MKMMKQFTGGGKGGK 440
Db | | : : | | |
415 MKALMKR-----KKGR 426

RESULT 13
Q8XF48 PRELIMINARY; PRT; 453 AA.
ID Q8XF48
AC Q8XF48;
```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 4.5S-RNP protein, GTP binding export factor, part of signal
DE recognition particle with 4.5 RNA (Signal recognition particle
DE protein).
DE FFH OR STM2677 OR STY2864.
GN OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OC NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE008821; AAL21566.1; -.
DR FGENES; AL627276; CAD05856.1; -.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54_1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 453 AA; 49799 MW; CF0F70616A2111BD CRC64;

Query Match 51.6%; Score 1173.5; DB 16; Length 453;
Best Local Similarity 51.4%; Pred. No. 3.7e-55;
Matches 233; Conservative 93; Mismatches 120; Indels 7; Gaps 2;

QY 3 FGLSERLQATMQKRGKGLKLTADIKIMREVLALFEADVNFVKYKFIKTSERALG 62
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 FDNLTDLRLSRLTNRISGRGLTDNVKLTLEVRMALLEADVALPVVREFINRVKEAVG 61

QY 63 SDVMGSLTPGQGVTKIVQDELTKMGENTSNINSNRPPTVVMVGLQAGKTTAGKLA 122
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 HEVNSLSLPGQGFYKIVSELSAAMGEENTNLNLAQPPAVLWMLGLOAGKTTSGKLG 121

QY 123 LLMKKYKKNKPELVAAIDYRPAINLQTVGKQIDIPVYSEGQVQKPOQIVTNALKHAK 182
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 KFLREKHKKVLVVSADVYRPAIKLETLAEQGVDFPSPYQKQPVDIVNAALKK 181

```

```

QY 183 EHLDFVIIDTAGRIHIDEALMNLKEVKEIAKPNELMVLVYDSMTGQDAYNVAESDDQLD 242
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 KEYDVLVLLTAGRLHVDEAMDEIKQVHASIKPVETLFFVDAMTGDAAATAKAFNEALP 241

QY 243 YTGVTILKLDGDRGGAALSIRSVTQKPIKFGVMSKLDGLSELFHFERVASRLGMDV 302
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 LTFGVVLTQVGDGARGGAALSIRHITGPKIFLGVGEKTDALFEFHDRIASRLGMDV 301

QY 303 SLIEKAQDVQDKAKADLEKMKRE--SSEFTLDDFLEQLDQVKNLGPLDDLMKMPGNMK 361
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 SLIEDIESKVDRAQAELKATLKKGGDFDLNDFLEQLKQKMKNGGMSLGMKLPQMG 361

QY 362 GLDKLNMSKQIDHIKAITQSMTPAERNPDTLVNVRKKRIAKSGRSLSQEVNRLMKOF 421
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 DNVKSQMDKVLVMEAIINSMTLKERAKPEIKSGRSKRRIAQCGMGQVQDVNRLKQ 421

QY 422 DMKMKMKOFTGGGKGGKKNQNWLMKMNLP 454
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 422 DMORMMK-----KMKKGGMAKMKRSMKGMPP 448

RESULT 14
Q9KUG1 PRELIMINARY; PRT; 461 AA.
AC Q9KUG1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Signal recognition particle protein.
GN VC0360.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004142; AAF93728.1; -.
DR HSP; O07347; IFFH.
DR TIGR; VC0560; -.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000897; SRP54.
DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00300; SRP54; 1.
KW Complete proteome.
SQ SEQUENCE 461 AA; 50373 MW; 616E9291972B3CDB CRC64;

Query Match 50.6%; Score 1150.5; DB 16; Length 461;
Best Local Similarity 50.8%; Pred. No. 6.4e-54;
Matches 228; Conservative 91; Mismatches 127; Indels 3; Gaps 2;

QY 3 FGLSERLQATMQKRGKGLKLTADIKIMREVLALFEADVNFVKYKFIKTSERALG 62
D 1: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 FDNLTDLRLSRLTNRISGRGLTDNVKLTLEVRMALLEADVALPVVREFINRVKEAVG 61

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 01:10:37 ; Search time 1563 Seconds
(without alignments)
4714.617 Million cell updates/sec

Title: US-09-943-108a-2
Perfect score: 2274
Sequence: 1 MAFGLSERLQATMQRKQK.....GKKGRNQNMGLKGNLPP 455

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlip

-MODEL=frame+p2n.model
-Q=/cgn2_1/USPTO_spool/US069943108/runat_21022003_151631_1824/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cgi -LIST=45
-DOCALLIGN=200 -THRESHOLD=100 -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USRP=US069943108 @CGN_1_1_2874 @runat_21022003_151631_1824 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRESH=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estio:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vit:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
c 1	1074	47.2	2169	17	BH770912	BH770912 LLMtag64
c 2	757.5	33.3	875	17	AF075878	AF075878 EST554007
c 3	715.5	31.5	773	13	BI934118	BI934118 EST554007
c 4	631	27.7	656	14	BQ805434	BQ805434 WHE3566_H
c 5	630	27.7	711	12	BF940797	BF940797 7D99a06_x
6	616.5	27.1	2387	11	AK011928	AK011928 Mus muscu
7	594.5	26.1	688	17	BH394904	BH394904 AG-ND-138
8	562.5	24.7	605	10	BE432304	BE432304 EST398833
9	558	24.5	593	10	AW648731	AW648731 EST327101
10	555.5	24.4	581	9	AI977909	AI977909 496035806
11	534	23.5	677	13	BM413418	BM413418 EST587745
12	520.5	22.9	1200	11	AI109321	AI109321 zea mays
13	513.5	22.6	533	10	AW930742	AW930742 EST356585
14	506.5	22.3	677	14	BQ996389	BQ996389 QGG12107
15	485	21.3	514	13	BM323434	BM323434 PIC1_19_C
16	479	21.1	1020	17	CNS0780K	AL437274 T3 end of
17	473	20.8	791	17	AQ327477	AQ327477 nxb0041K
18	465.5	20.5	1032	11	AY106667	AY106667 zea mays
19	463.5	20.4	483	10	BE596266	BE596266 P11_51_F0
20	451	19.8	777	17	BH402176	BH402176 AG-ND-122
21	448.5	19.7	490	13	BM325805	BM325805 PIC1_53_A
22	438.5	19.3	554	17	BH369841	BH369841 AG-ND-126
23	437.5	19.2	481	17	BH376496	BH376496 AG-ND-178
24	433.5	19.1	880	17	CNS077VW	AL433218 T7 end of
25	420	18.5	465	9	AI087737	AI087737 SWAMCAC23
26	419	18.4	476	10	AW033339	AW033339 EST276910
27	419	18.4	899	14	BQ736950	BQ736950 AGENCOURT
28	414.5	18.2	547	10	BE461473	BE461473 EST412892
29	412.5	18.1	688	13	BI933601	BI933601 EST553490
30	412.5	18.1	1108	9	AU091185	AU091185 AU091185
31	408	17.9	1059	14	BM806783	BM806783 AGENCOURT
32	405	17.8	1012	17	AF094895	AF094895 AF094895
33	400	17.6	949	17	CNS076TA	AL431828 T7 end of
34	398.5	17.5	814	17	AZ529841	AZ529841 ENT0084TR
35	392.5	17.3	889	17	AZ682189	AZ682189 ENT0084TR
36	389.5	17.1	773	13	BI454259	BI454259 603170687
37	388	17.1	761	9	AL655153	AL655153 AL655153
38	385.5	17.0	656	12	BF641333	BF641333 NF064F12I
39	380.5	16.7	694	13	BJ473333	BJ473333 BJ473333
40	378.5	16.6	874	12	BG679799	BG679799 602626480
41	377	16.6	707	17	AQ654845	AQ654845 Sheared D
42	375.5	16.5	753	13	BG921695	BG921695 602825514
43	375	16.5	382	10	AW649149	AW649149 EST327603
44	374.5	16.5	975	13	BG919869	BG919869 602823428
45	369.5	16.2	573	12	BG317280	BG317280 947027E11

ALIGNMENTS

RESULT 1
BH770912/c
LOCUS BH770912
DEFINITION LLMtag644 MG1363 Random Sequence Tag Library Lactococcus lactis subsp. cremoris genomic, DNA sequence.
ACCESSION BH770912
VERSION BH770912.1
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris.
ORGANISM Lactococcus lactis subsp. cremoris.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
REFERENCE 1 (bases 1 to 2169)
AUTHORS Bolotin,A., Ehrlich,S.D. and Sorokin,A.
TITLE Studies of genomes of dairy bacteria Lactococcus lactis

JOURNAL
COMMENT

Sci. Aliments, (2002) In press
Contact: Sorokin A
Genetique Microbienne
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL1403 is ffh (91%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 2141.
Location/Qualifiers
1. .2169
/organism="Lactococcus lactis subsp. cremoris"
/strain="MGI363"
/db_xref="taxon:1359"
/clone_lib="MGI363 Random Sequence Tag Library"
/notes="Vector: pSGM02; Site_1: SmaI; Library of
chromosomal fragments of L.lactis strain MGI363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 642 a 429 c 380 g 718 t
ORIGIN

FEATURES
source

Alignment Scores:
Pred. No.: 4.02e-108 Length: 2169
Score: 1074.00 Matches: 218
Percent Similarity: 73.86% Conservative: 73
Best Local Similarity: 55.33% Mismatches: 82
Query Match: 47.23% Indels: 21
DB: 17 Gaps: 2

US-09-943-108A-2 (1-455) x BH770912 (1-2169)

QY 76 IleLysIleValGlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsn 95
Db 2168 ATCAAAATGTGTGAAGAATTAACAATATCTATGCGCGGTGTAAGCCCAATATCTC 2109
QY 96 MetSerAsnLysProThrValValMetMet-VaIGlyLeuGlnGlyAlaGlyLysTh 115
Db 2108 AAATCTCTAAATCCCAACAATATCATGATGGTGGTTTTCAGAGGGCAGGTAAAC 2049
QY 115 rThrThrAlaGlyLysLeuAlaLeuLeuMetArGLysLysrAsnLysPrometLe 135
Db 2048 AACCTTGTCTGTGAATCTGCAAAAACTTAAGAAGAACAAATGCTGCCACTGAT 1989
QY 135 uVal-AlaAlaAspIleTyArgProAlaAlaIle-AsnGlnLeuGlnThrValGlyLys 154
Db 1988 GATTGGCAGCGAGCTTATPCGTCAGAGCGGATGGACCAATTAACCACTTGGAGAA 1929
QY 155 GlnIleAspIleProValTySerGluGlyAspGlnValLysProGlnGlnIleValThr 174
Db 1928 CAACCTGACATCTCTGTATTATGATGAAGACGCTCGAAGAACCCGTTAATATGTCGT 1869
QY 175 AsnAlaLeuLysHisAlaLysGluGlnHisLeuAspPheValIleleAspThrAlaGly 194
Db 1868 AATGTCTCTTTAAAGCTCAAGAAGACGTAAGAATATGTTGTTGATACGCGAGGT 1809
QY 195 ArgLeuHisIleAspGluAlaLeuMetAsnGlnLeuLysGluValLysGluIleAlaLys 214
Db 1808 CGCTGTGAATGTGATACATATGATGATGATGATGATGATGATGATGATGATGATGAT 1749
QY 215 ProAsnGlnIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAla 234
Db 1748 CCAACAGAAATTTACTCGTTGTGATGATGATGATGATGATGATGATGATGATGAT 1889
QY 235 GluSerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAsp 254
Db 1688 AAAAATTTTGTGATGAAAAATTAGATATACCGGTGTATCATCAACCAACTCGATGGGAC 1629
QY 255 ThrArgGlyGlyAlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheVal 274
Db 1628 ACACGTGTGTGGCGTTTATCAATTCGTGAATATTCGGAATATACCGGAACCTTATAATTTACA 1569

QY 275 GlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArg 294
Db 1568 GGGATGCTGTAAGAAATTAACCGACCTTGAACCTTTCTACCTGATGATGATGATGATGAT 1509
QY 295 IleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAspValAspGln 314
Db 1508 ATTCTGTGTATGGTGATATGCTCACTGCTGATTGAGAAGCGCAAGCTAATATTGATGAG 1449
QY 315 GluLysAlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLeuAspPhe 334
Db 1448 GAACAATCTGCAAACTCGCTGAAATAATGGCTGATAACCCGTTTTCATTATGAGATTTC 1389
QY 335 LeuGluGlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIle 354
Db 1388 GTTGAGCACTAGATCAAGTTACATATGAGGACCAATGGAAGATATTATGAAATGATT 1329
QY 355 ProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAsp 374
Db 1328 CCAGGAATGTCACAAATGCTGCTGGTAAAGTAAAGTTGACCTTAAGATGTTGCT 1269
QY 375 HisIleLysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeu 394
Db 1268 AGAAACGAGCAATGTTATGTCATGACCCGCTGACGTCATTTAGAACGCAACTA 1209
QY 395 AsnValSerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGlnGlnValAsn 414
Db 1208 TCGCCAGCTGCTGCTGCGAATCGCCGAGGTCTCGGAATTCATTTATTGAATTAAT 1149
QY 415 ArgLeuMetLysGlnPheAsn-----Asp 422
Db 1148 AAATTCATTAGCAATTTAATCAATCAACAAAGAAATGATCAAGGAATCATGAATGTTGAT 1089
QY 423 MetLysLysMetLysGlnPheThrGlyGlyGlyLysLysLysLysLysLysLys 442
Db 1088 ATGAATGCGATGATGACAAATAATGATGGGTGAGCTGCTGGA----- 1047
QY 443 GlnMetGlnAsnMetLeuLysGlyMetAsnLeuPro 454
Db 1046 CAAATGCCAATATGCCACCGGTTCTGGAATGCCA 1011
RESULT 2
LOCUS AF075878 875 bp DNA linear GSS 29-AUG-2000
DEFINITION AF075878 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 245-T3, DNA sequence.
ACCESSION AF075878
VERSION AF075878.1 GI:3320748
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE 1 (bases 1 to 875)
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
comparison to the Escherichia coli K12 genome
JOURNAL FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE 99243757
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1. .875
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="245-T3"
/note="Vector: Lambda DASH II; sequenced using LI-Cor

```
BASE COUNT      178 a      266 c      217 g      214 t
ORIGIN
Alignment Scores:      Length:      875
Pred. No.:      757.50
Score:      Matches:      152
Percent Similarity:      73.08%
Conservative:      57
Best Local Similarity:      53.15%
Mismatch:      76
Query Match:      33.31%
Indels:      2
DB:      17
Gaps:      1

US-09-943-108a-2 (1-455) x AF075878 (1-875)
QY 87 MetGlycylGluAsnThrSerIleAsnMetSerAsnLysProProThrValValMetMet 106
Db 867 ATGGCGAGAGAACACACAGCTGAATTTTAC-TGCACGCCGACGCGTAGTATTGATG 809
QY 107 ValGlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArg 126
Db 808 CGGGTCTCGAGGGGCGGTAACACACAGCTGCGTAAATTTGTTGGTAAATTTCTTGGC 749
QY 127 LysLysThrAsnLysLysProMetLeuValAlaAlaAspIleTyArgProAlaAlaIle 146
Db 748 GAGAGACACAGAGAAAGTCTGCTGCTCTGCGACGCTATATGCGCGGCGGCGATC 689
QY 147 AsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTySerGluGlyAspGln 166
Db 688 AATACGCTCGAATACCTGGCTGACAGAGTGGCGGTGATTTCTCCGCTGTGATGTCGC 629
QY 167 ValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHisLeuAsp 186
Db 628 CAGAAACCGGTTGATATCGTCAACGCCGCGCTGAAAGAGCGAAGCTCAATCTACGAC 569
QY 187 PheValIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeu 206
Db 568 GTCTGCTGCTGATACCGCGGCTGCTGCTGATGTTGACGAGCGGATGATGAGCAATC 509
QY 207 LysGluValLysGlnIleAlaLysProAsnGluIleMetLeuValValAspSerMetThr 226
Db 508 AATACGCTCCAGCTTCTATCAACCCAGTAGAAGACGCTTTGTCGTGATGATGAGCAC 449
QY 227 GlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyVal 246
Db 448 GTCAGATGCGCGGAGTAACCGCAAGAGCTTTAAGAGAGCGCTGCCCTTAACGCGGCG 389
QY 247 ThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerVal 266
Db 388 GTGCTGACCAAGTTGATGTTGATGCGCGTGGCGGTGCGGCGCTCTCTATTGTCATATC 329
QY 267 ThrGlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPhe 286
Db 328 ACCGGAAGCCGATTAATTTCTCGGTGTCGGGAGAAACGAGCGCTGGAGCCATTC 269
QY 287 HisProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGlu 306
Db 268 CATCCGGATCGTATCGCTCCGCTATCTCGGATGCGGCGACGCTACTGCTCTTATCGAA 209
QY 307 LysAlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGlu 326
Db 208 GATATCGAAAGCAAGTTGATGCGGCGACAGGCTGAGAGCTGGGACCAACTGAAGAA 149
QY 327 ---SerSerPheThrLeuAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
Db 148 GCGAGCGGTTTCGACTCGAAGCTTCTCGAAGACAGCTCAACACAGATGAAGAAACATGGG 89
QY 346 ProLeuAspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLys 365
Db 88 GGTATGCCACGCTGATGGGCAATATACCGGCGATGGGCGAGATTCGGGACACAGTTAA 29
QY 366 LeuAsnMetSerGluLys 371
Db 28 TCACAGATGATCACAG 11
```

RESULT 3

BI934118

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BI934118 773 bp mRNA linear EST 18-OCT-2001
clone cTOD18J16 5' end, mRNA sequence.

BI934118 GI:16248590

EST.

tomato.

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 773)

REFERENCE

AUTHORS

van der Hoeven,R.S., Bezzerides,J.L., Karanycheva,S.A., Tsai,J.,
Unterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, anthesis (2001)

Unpublished (2001)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

FEATURES

source

1..773
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD18J16"
/tissue.lib="tomato flower, anthesis"
/tissue.type="flower"
/dev_stage="anthesis"

/note="vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; supplier: Cornell University; sequencing: The
Institute for Genomic Research; Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 5,83e-69 Length: 773

Score: 715.50 Matches: 134

Percent Similarity: 75.19% Conservative: 60

Best Local Similarity: 51.94% Mismatches: 63

Query Match: 31.46% Indels: 1

DB: 13 Gaps: 1

US-09-943-108a-2 (1-455) x BI934118 (1-773)

QY

39 LeuPheGluAlaAspValAsnPheLysValValLysGluPheIleLysThrValSerGlu 58

Db 1 CTTTATGAGCTGATGTAGTCTCCAGTGTTCAGAGGTTGTTTCAGCTGTTAGTGAA 60

QY

59 ArgAlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIle 78

Db 61 GAAGCGCTGGCACCTGGTTGATTCAGAGGTGTAAACACAGATTCAGACTAGTTAAGACT 120

QY

79 ValGlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsn 98

Db 121 GTACGTGACGACCTTGTGAACACTGATGGGTGGAGAGGTTCTTGAACTGGTTTTCGTAAA 180

QY

99 LysProThrValValMetMetValGlyLeuGlnGlnValAlaGlyLysThrThrAla 118

Db 181 TCATAGCCCAACCGCTAATCTATTGGCCGGTCTACAGGTTGTGGGAGACACACTGTAGT 240

```

QY 119 GlyLysLeuAlaLeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAla 138
Db 241 GCAGAGTGTATATCTA---AGAGAGCGGTAGAGTTCATGCTGATTGCTGGA 297

QY 139 AspLeuTyrArgProAlaAlaLeuAsnGlnLeuGlnThrValGlyLysGlnLeuAspLeu 158
Db 298 GACGTGTACAGACCTGCTGCTATTGACCACTGTATTATTGGTGAACAGGTTGATGTA 357

QY 159 ProValTyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLys 178
Db 358 CTTGTTTATGACGAGACAGATGTAAACCTGCGAATAGCCGACACAGGATACAA 417

QY 179 HisAlaLysGluGlnHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIle 198
Db 418 GAGGCCAAAGAAAGATGTAGATGTAGTCATATGATGATACAGCTGGAAGACTTCAGATA 477

QY 199 AspGluAlaLeuMetAsnGlnLeuLysGluValLysGluIleLeuAlaLysProAsnGluIle 218
Db 478 GATAAACTATGATGGATGAATTAAGACGCTGAACGGGTACTGAACCCACACAGAGTT 537

QY 219 MetLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAsp 238
Db 538 TTTACTTTGTGTGATGATGATGCTGCGCAGAGCTGCAGCTTGTGTCACACATCAAT 597

QY 239 AspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyIle 258
Db 598 CTCGAATTTGGATTTACTGTGCTCCTTTGACGAGCTAGATGAGATCTAGGGGTGGA 657

QY 259 AlaAlaLeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlu 278
Db 658 GCAGCTTTTGTGTGTCAGAGGAGTATCAGGAAGCCCAATCAAGCTTCGTGAGAGGGGTGAA 717

QY 279 LysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeu 296
Db 718 CGATGGAGGACCTTGACCTTCTATCTGACCGCATGGCTGGACGATTTTGA 771

RESULT 4
BQ805434
LOCUS BQ805434 656 bp mRNA linear EST 31-JUL-2002
DEFINITION WHS3566_H12_P242S Wheat developing grains cDNA library Triticum
aestivum cDNA clone WHE3566_H12_P24, mRNA sequence.
ACCESSION BQ805434
VERSION BQ805434.1 GI:22029643
SOURCE EST.
ORGANISM bread wheat.
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 656)
Altenbach,S., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
The structure and function of the expressed portion of the wheat
genomes - Developing grains cDNA library
Unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105593773
Fax: 5105593818
Email: oanderson@pw.usda.gov
Sequences have been trimmed
Quality sequence with phred score less than 20
Seq primer: SK primer.
Location/Qualifiers
1..656
/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3566_H12_P24"
/clone_lib="Wheat developing grains cDNA library"
FEATURES
source

```

```

/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda ZAP II, excised phagemid; Site_1:
ECORI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24
32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24,
28, 32, 34 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16
, 20, 24, 28, 30 DPA and total RNA was prepared by S.
Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give pBluescript SK(-) phagemids
in the TJ Clone lab (Chin, close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
BASE COUNT 200 a 115 c 179 g 162 t
ORIGIN

Alignment Scores:
Pred. No.: 11e-59 Length: 656
Score: 631.00 Matches: 116
Percent Similarity: 75.23% Conservative: 48
Best Local Similarity: 53.21% Mismatches: 54
Query Match: 27.75% Indels: 0
DB: 14 Gaps: 0

US-09-943-108A-2 (1-455) x BQ805434 (1-656)

QY 141 TyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnLeuAspIleProVal 160
Db 3 TACAGGCTGCTGCCATGTATCACTACTGTACTGGTGAACAGGTTGGTGCAGAT 62

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
Db 63 TACTCAGAAGGACCGCGGCCAAACCTGCAGAAATPAACCAAGATCCGTGGAAGAGCG 122

QY 181 LysGluGlnHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
Db 123 AAAAGAAATATATTGATGCGATCGTAGTGATATCTGCTGCAGACTGCAGATTGATATA 182

QY 201 AlaLeuMetAsnGlnLeuLysGlnValLysGlnLeuAlaLysProAsnGlnIleMetLeu 220
Db 183 ACATGATGTTGATTAATGAAGAGATGAAGAGGCGACGTATTCTACAGATTGGTT 242

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
Db 243 GTTGTGTATGTCATGACTGCGCCAGGAAGCTGCGAGCTTGGTCAACCTTCAATTTGAA 302

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyClyAlaAla 260
Db 303 ATTGTATATCCGTGCTATATTGACTAAATGATGATGATTCAGAGGGTGGAGACGA 362

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGlnLysLeu 280
Db 363 CTATGTCGAAGAGGTCCTCGAAGACCCATCAAGTTTGTGGCGAGAGAGCGAATG 422

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300

```


REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
11042159

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishigami, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11076861

REFERENCE

AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauber, J., Suzuki, R., Tomita, M., Wagner, B., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kaniya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
PUBMED
11217851

REFERENCE

AUTHORS

5 (bases 1 to 2387)
Aachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Ouackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
Location/Qualifiers
1..2387
/organism="Mus musculus"

REFERENCE

AUTHORS

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
Location/Qualifiers
1..2387
/organism="Mus musculus"

gene

CDS

/strain="C57BL/6J"
/db_xref="PANTOM_DB:2610209C12"
/db_xref="MGDI:1905318"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
1..2387
/gene="Srp54"
238..1752
/gene="Srp54"
/note="data source:MGD, source key:MGDI:1345087,
evidence:ISS
putative
signal recognition particle 54 kDa"
/codon_start=1
/protein_id="BAB27921.1"
/db_xref="GI:12848354"
/db_xref="MGDI:1346087"
/translation="MVLADLGRKTSALRSLSNATINEEVLNMLKEVCTALLADV
NKLKVKQRENVKSAIDLEMASGLNKRKMQHAFVKELKLVDPGVKATPTKQKN
VIMFVGLOGSGKTTCSKLAIYVYORKMKTCLICADTFRAGAEQDKQKATKARIPFY
GSYTFMDSVIIASSEGVKEFKNFENFELIIVDSGRHQEDSLFEMIQVNATQPIFY
YVMDASIQACEAQAKAFKDVASIVITKLDHGKGGALSAVAATKSPFIIFGTG
EHIDDFEFKTPFTSKLGMGDIKLVNELKDDNHALTEKLKHGQDLTQWYE
OFONTMKGPPSOILGMIKPGSGTDFMSKNEQSMARLKKLMTIMPMDQELDSTDG
AKVESKOPGRQVRGSGVSPEDYQELLTQYTFKAQWVKMGIGIKLFGKGGDMKNV
SSQWAKUNQQAAMMDPRVLIHMGMGAGLQSMRQFQGAAGNKGMMGFNM"
BASE COUNT 741 a 420 c 560 g 666 t
ORIGIN
Alignment Scores:
Pred. No.: 2,29e-57 Length: 2387
Score: 616.50 Matches: 150
Percent Similarity: 52.12% Conservative: 96
Best Local Similarity: 31.78% Mismatches: 201
Query Match: 27.11% Indels: 25
DB: 11 Gaps: 8
US-09-943-108a-2 (1-455) x AK011928 (1-2387)

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

```

QY 161 TyrSerGluGlyAspGlnValLysProGlnGlnLeuValThrAsnAlaLeuLysHisAla 180
    ||| : : : : : ||| : : : : :
Db 715 TATGACAGCTACTAGAAATGGATCCTGTCATCATCTCTTGAAGAGTGGAGAAATTC 774

QY 181 LysGluGluHisLeuAspPheValIleLeuAspThrAlaGlyArgLeuHisIleAspGlu 200
    ||| : : : : : ||| : : : : :
Db 775 AAAATGAAAAATTTTGAATATATATTTGTTATCAAGTGTGTCACAAACAAAGAAC 834

QY 201 AlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeu 220
    ||| : : : : : ||| : : : : :
Db 835 TCTTTATTTGAAGAAATGCTTCAAGTTCTTAAGCTATACACTATACATCTTTAT 894

QY 221 ValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspGln 240
    ||| : : : : : ||| : : : : :
Db 895 GTGATGATGATCCATCGACAGGCTTGTGAGGCCCGAGGCGCTTTAAGACAA 954

QY 241 LeuAspValThrGlyValThrLeuThrLysLeuAspGlyAspThrArgGlyAlaAla 260
    ||| : : : : : ||| : : : : :
Db 955 GTAGATGTAGTTCAGTAATAGTGACAAACTCGACGTCATGCAAGAGGCGGTGCT 1014

QY 261 LeuSerIleArgSerValThrGlnLysProIleLysPheValGlyMetSerGluLysLeu 280
    ||| : : : : : ||| : : : : :
Db 1015 CTATGTCAGTGTGCCCAAAAGTCCATTTATTTCAATGTGACAGGGAACATATA 1074

QY 281 AspGlyLeuGluLeuPheHisProGluArgMetAlaSerArgIleLeuGlyMetGlyAsp 300
    ||| : : : : : ||| : : : : :
Db 1075 GATGATTTTGAACCTTCAAAACACACAACTTTCATCAGCAAACTCCTTGGATGGTGT 1134

QY 301 ValLeuSerLeuIleGluLysAlaGlnGln---AspValAspGlnGlnLysAlaLysAsp 319
    ||| : : : : : ||| : : : : :
Db 1135 ATTGAAGGACTGATTTGATAAAGTCAATGAATTAAGTTGATGATGATGATGATGAT 1188

QY 320 LeuGluLysLysMetArgGluSerSerPheThrLeuAspAspPheLeuGluGlnLeuAsp 339
    ||| : : : : : ||| : : : : :
Db 1189 CTTATAGAGAGTTGAGACGCTGAGTTACATTTGCGACATGATGATGATGATGATGAT 1248

QY 340 GlnValLysAsnLeuGlyProLeuAspAspIleMetLysMetIleProGlyMetAsn--- 358
    ||| : : : : : ||| : : : : :
Db 1249 AATATTGAAATGGGCCCATTCAGTCAGATATTGGGGATGATCTCTGGCTTGGCACA 1308

QY 359 -----LysMetLysGlyLeuAspLysLeuAsnMetSerGluLysGlnIleAspHisIle 376
    ||| : : : : : ||| : : : : :
Db 1309 GATTTTATGACAAAGGAATGACAGAGTCAATGGCA-----AGGCTGAAGAACTG 1362

QY 377 LysAlaIleIleGlnSerMetThrProAlaGluArgAsnAsnProAspThrLeuAsnVal 396
    ||| : : : : : ||| : : : : :
Db 1363 ATGCAATCATGGACATGATGACGATCAAGAACTGACAGTACATGATGATGATGATGAT 1422

QY 397 -----SerArgLysLysArgIleAlaLysGlySerGlyArgSerLeuGln 411
    ||| : : : : : ||| : : : : :
Db 1423 TTCAGTAAGCAACACGAGGAGATCCAAAGAGTTGCCCGGGATCAGGTGTGTCAACAAGA 1482

QY 412 GluValAsnArgLeuMetLysGlnPheAsnAspMetLysLysMetLysGln----- 429
    ||| : : : : : ||| : : : : :
Db 1483 GATGTTCAAGAAGCTCTGACCCAGTATACCAAGTTTGCACAGATGTCGTAAGAAGTGGGA 1542

QY 430 -----PheThrGlyGlyGlyLysGlyLysGlyLysGlyLysArgAsnGlnMet 444
    ||| : : : : : ||| : : : : :
Db 1543 GGTATCAAGAGCACTTTTCAAGCGGTGATATGCTATAGATGTGAGTCAATGATGATGAT 1602

QY 445 -----GlnAsnMetLeuLysGlyMetAsn 452
    ||| : : : : : ||| : : : : :
Db 1603 GCAAAATTAACCAACAATGGCCAAATGATGGAC 1638

RESULT 7
BH394904
LOCUS BH394904
DEFINITION AG-ND-138B11.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138B11
, DNA sequence.
ACCESSION BH394904
VERSION BH394904.1 GI:17341045
KEYWORDS GSS.
SOURCE African malaria mosquito.

```

```

ORGANISM Anopheles gambiae
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae,
          Anopheles.
REFERENCE 1 (bases 1 to 688)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-138B11.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: B.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seg primer: M13 For
Class: BAC ends.
          Location/Qualifiers
            source
              1..688
                /organism="Anopheles gambiae"
                /strain="PEST"
                /db_xref="taxon:7165"
                /clone="AG-ND-138B11"
                /clone_lib="ND-TAM"
                /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 223 a 118 c 154 g 193 t
ORIGIN

Alignment Scores:
Pred. No.: 1,29e-55 Length: 688
Score: 594.50 Matches: 129
Percent Similarity: 73.39% Conservative: 42
Best Local Similarity: 55.36% Mismatches: 55
Query Match: 26.14% Indels: 7
DB: 17 Gaps: 1

US-09-943-108a-2 (1-455) x BH394904 (1-688)

QY 60 AlaLeuGlySerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleVal 79
    ||| : : : : : ||| : : : : :
Db 2 GCTTTAGGACAAATGTTTAAACATCCATTACCCGGGACAAATGATGACAAAAATGTT 61

QY 80 GlnAspGluLeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLys 99
    ||| : : : : : ||| : : : : :
Db 62 CACGATGATGTAGTGGATTTAATGGGAGGAACCAATGAAGGCTTAATCTTCCGAAAAA 121

QY 100 ProThrValValMetValGlyLeuGlnGlyAlaGlyLysThrThrAlaGly 119
    ||| : : : : : ||| : : : : :
Db 122 ---CCACATTTATCTAGTGTGAGTTTACAGGTTCTGGAATCTATCTCTCCGGG 178

QY 120 LysLeuAlaLeuLeuMetArgLysLysTyrAsnLysLysPrometLeuValAlaLys 139
    ||| : : : : : ||| : : : : :
Db 179 AATTTAGCAAACTTCTCTGAAGAGAAAAAGAGCAAAATCGTTATTGTTAGCATCTGAC 238

QY 140 IleTyrArgProAlaAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIlePro 159
    ||| : : : : : ||| : : : : :
Db 239 GTTTATCTCTCTGCTGCTGATTCAGCTAAAAGTATTAGTAGTCAGACAGGGTCTCT 298

QY 160 ValTyrSerGluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHis 179
    ||| : : : : : ||| : : : : :
Db 299 GTATATCTAGTGAAGGAGTAACAAATCGGTTTCAGATTCTCAGAAATCAATAGAAATTT 358

QY 180 AlaLysGluGluHisLeuAspPheValIleIleAspThrAla-GlyArgLeuHisIleAs 199

```

```

|||||:  ||| :|||:|||||:|||||:|||||:|||||:|||||
Db 359 GCAAAGCAAAATAAGCACGATGTTATCATATAGATACCGCAGGGCGTTTGGCTATCGA 418

QY 199 pGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlu-Ilex 219
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db 419 TGAAGAGATGATGAACGAAATTCCTAACGTTCCACGTCGTCTGTAACCTACAGAACTCT 478

QY 219 etLeuValValAspSerMetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspA 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 479 TTTTCGTAGTTGACTATGACTGGGAGGATGCTGTGATCCAGCAAAAGCTTTTAACG 538

QY 239 spGlnLeuAsp-ValThrGlyValThrLeuThrLysLeuAspGlyAspThrArg-GlyEl 258
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db 539 ATGTCCTTAATATACGGGTGATGTTCTTACTAAATTAGATGGTGATACGCGAGGTGG 598

QY 258 yAlaAlaLeuSerIleArgSer-ValThrGlnLysProIleLysPheVal-GlyMetSer 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 599 GCGCGCATAAACATCCGTTCCGTTAGTAGAAAAACCAATTAAGTTAATTCTTACCGGT 658

QY 278 GlnLysLeuAspGlyLeuGluLeuPhe 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 659 GAGAAATGATGCTCTGGGATATTTT 685

RESULT 8
BE432304
LOCUS
DEFINITION
EST398833 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG7M2, mRNA sequence.
ACCESSION
VERSION
BE432304.1 GI:9430147
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1. (bases 1 to 605)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
location/Qualifiers
1..605
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG7M2"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT
190 a 97 c 159 g 159 t
ORIGIN
Alignment Scores:
Pred. No.: 3,86e-52 Length: 605
Score: 562.50 Matches: 105
Percent Similarity: 75.25% Conservative: 47
Best Local Similarity: 51.98% Mismatches: 49

```

```

Query Match: 24.74% Indels: 1
DB: 10 Gaps: 1
US-09-943-108A-2 (1-455) x BE432304 (1-605)
QY 65 ValMetGlnSerLeuThrProGlyGlnGlnValIleLysIleValGlnAspGluLeuThr 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2 TTGATTGCGAGGTGTAAACCAGATCAGCAACTAGTTAAGACTGTACGTGACGAGCTGTG 61

QY 85 LysLeuMetGlyClyGluAsnThrSerIleAsnMetSerAsnLysProProThrValVal 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 62 AAATGATGGTGGAGAGGTTTCTGAACTGGTTTGGTAAATCTAAGCCCAACCGTAATA 121

QY 105 MetMetValGluGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeu 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 122 CTATTGCCCGGCTCAGAGGTGTGGGAAGACAACCTGTAGTGCAAGCTTACCTTATAT 181

QY 125 MetArgLysLysTyrrAsnLysLysProMetLeuValAlaAlaAspIleTyrrArgProAla 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 182 CTA--AAGAAGCAGGGTAAGAGTTGCATGCTGATGCTGGAGAGCTGTACAGACCTGCT 238

QY 145 AlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrrSerGluGly 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 239 GCTATTGACCAACTGTTTATTGTTGGTAAACAGGTTGATGATGCTGTTTATGCGACGAGA 298

QY 165 AspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluGluHis 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 299 ACAGATGTAAACCTGCAGAAATAGCCCGACAGGATTCACAGAGCCCAAAAAGAAAT 358

QY 185 LeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsn 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 359 GTAGATGTAGTCATATGATGATACAGCTGGAGAGCTTCAGATAGATAAAACTATGATGAT 418

QY 205 GluLeuLysGluValLysGluIleAlaLysProAsnGluIleMetLeuValValAspSer 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 419 GAATTAAGACGCTGAACGGGTACTGAACCCACAGAGGTTTACTTGTGTGGATGCA 478

QY 225 MetThrGlyGlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThr 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 479 ATGACTGGCCAAAGACTCGACGTTTGTGTCAACATCATCAATCTCGAAATGGGAATACT 538

QY 245 GlyValThrLeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArg 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 539 GGTCCCATCTTGGACGAGCTAGATGGAGATCTAGSGGTGGAGCAGCTTAAAGTGCAAA 598

QY 265 SerVal 266
|||
Db 599 GAGGTA 604

RESULT 9
AW648731
LOCUS
DEFINITION
EST327101 tomato germinating seedlings, TAMU Lycopersicon
esculentum cDNA clone cLEI522 5', mRNA sequence.
ACCESSION
AW648731
VERSION
AW648731.1 GI:7409885
KEYWORDS
EST.
SOURCE
tomato.
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE
1. (bases 1 to 593)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
,S.D.
Generation of ESTs from germinating tomato seed
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University

```



```

:::|||||
Db 361 GTACCAGTTTACTCAGAGGAAGTCAAGCAAAACCTTCACAATATAGCCAAAAGGGTTG 420
:::|||||
QY 178 LysHisAlaLysGluLHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHis 197
:::|||||
Db 421 AAGGAGCAAAAGTCAATAGCGTGTATATATAGTGACACCGCTGGGAAGACTGCAG 480
:::|||||
QY 198 IleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAlaLysProAsnGlu 217
:::|||||
Db 481 GTAGATAAACAATGATGATGAGTGTGAAGAGAGTAAAAAAGCAGTGAATCCCTACAGAA 540
:::|||||
QY 218 IleMetLeuValLysSerMetThrGlyClnAspAla 230
:::|||||
Db 541 GTTCTTCTGTTGATGCCATGACTGGCCCAAGAAGCT 579

RESULT 11
BM4113418
LOCUS EST587745 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION CLEGG63C14 5' end. mRNA sequence.
ACCESSION BM4113418
VERSION BM4113418.1 GI:18265048
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 677)
AUTHORS Alcalá, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai
, J., Bougria, O., Kirchner, E., Utterback, T., Van Aken, S., Ronning
, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
JOURNAL
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
Location/Qualifiers
FEATURES
source
1..677
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG63C14"
/clone_lib="tomato breaker fruit"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCquadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 222 a 119 c 165 g 171 t
ORIGIN

Alignment Scores:
Pred. No.: 6,42e-49 Length: 677
Score: 534.00 Matches: 101
Percent Similarity: 75.38% Conservative: 46
Best Local Similarity: 51.79% Mismatches: 47
Query Match: 23,48% Indels: 1
DB: 13 Gaps: 0

US-09-943-108a-2 (1-455) x BM4113418 (1-677)

```

```

QY 152 ValGlyLysGlnIleAspIleProValItySerGluGlyAspGlnValLysProGlnGln 171
:::|||||
Db 92 GTGGGGGAACAGGTGTGATGCTACCTGTTTATGCAGCAGGAACAGATGTTAAACCTGCAGAA 151
:::|||||
QY 172 IleValThrAsnAlaLeuLysHisAlaLysGluLHisLeuAspPheValIleIleAsp 191
:::|||||
Db 152 ATAGCCGACAGAGATTACAGAGGCCCAAGAAAGAAATGTAGATGTAGTCATAATGAT 211
:::|||||
QY 152 ThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlnLeuLysGluValLysGlu 211
:::|||||
Db 212 ACATGCGAAGACTTCAGATAGATAAACTATGATGATGAATTTAAAGACGTGAACGG 271
:::|||||
QY 212 IleAlaLysProAsnGluIleMetLeuValLysSerMetThrGlyClnAspAlaVal 231
:::|||||
Db 272 GTACTGAACCCACAGAGAGTTTACTGTTGTGGATGCAATGACTGGCCCAAGAGCTGCA 331
:::|||||
QY 232 AsnValAlaGluSerPheAspGlnLeuAspValThrGlyValThrLeuThrLysGlu 251
:::|||||
Db 332 GCTTTGTGTCACAACTCAATTCGAATTTGAATTACTGGTGGCATCTTGACGAGCTA 391
:::|||||
QY 252 AspGlyAspThrArgGlyGlyAlaLeuSerIleArgSerValThrGluLysProIle 271
:::|||||
Db 392 GATGAGATTTCTAGGGGTGGAGCAGCTTTAGTGTCAAGGAGGTATCAGGAAGCCATC 451
:::|||||
QY 272 LysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMet 291
:::|||||
Db 452 AAGCTCGTAAGAAGGGGTGAAGTATGAGGAGCCTTGAACTTTCTATCTGACCGCATG 511
:::|||||
QY 292 AlaSerArgIleLeuGly-MetGlyAspValLeuSerLeuIleGluLysAlaGlnGlnAs 311
:::|||||
Db 512 GCTGGAGCTATTGTGAATAATGGGAGATGTTCTATGCTTTGTGAGAAAGCCCAAGAAAG 571
:::|||||
QY 311 pValGlnGlnLysAlaLysAspLeuGluLysLysMetArgGluSerSerPheThrLe 331
:::|||||
Db 572 TATGCTCAGAGAGATGCTGAGAAATTCAGAGAGATCATCAGATGTCGAAATTTGATT 631
:::|||||
QY 331 uAspAspPheLeuGluGlnLeuAspGlnValLysAsnLeuGly 345
:::|||||
Db 632 CAATGACTTCTGAGCAAACTCCGGCAGTTGCTAGGATGGGT 674

RESULT 12
AY109321
LOCUS Zea mays PCC070230 mRNA sequence.
DEFINITION Zea mays PCC070230 mRNA sequence.
ACCESSION AY109321
VERSION AY109321.1 GI:21212876
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1200)
AUTHORS Coe, E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
Location/Qualifiers
SOURCE
1..1200
/organism="Zea mays"
/db_xref="MaizeDB:633818"
/db_xref="taxon:4577"
/clone="PCC070230"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public

```

contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 387 a 216 c 317 g 279 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 4.16e-47 Length: 1200
Score: 520.50 Matches: 102
Percent Similarity: 70.82% Conservative: 63
Best Local Similarity: 43.78% Mismatches: 65
Query Match: 22.89% Indels: 3
DB: 11 Gaps: 2

US-09-943-108a-2 (1-455) x AY109321 (1-1200)

QY 208 GlnValLysGluLeuAlaLysProAsnGluLeuMetLeuValValAspSerMetThrGly 227
DB 11 GAAGTGAAGAGGCTGTTAACTCCACAGAAAGTTCTGCTGTGATGAGCCAGTCTGC 70

QY 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247
DB 71 CRAAGAGCAGCAGCTAGTCAACACCTTCAATATTGAGATTGGTATTAACCTGGTGCATA 130

QY 248 LeuThrLysLeuAspGlyAspThrArgGlyGlyAlaAlaLeuSerIleArgSerValThr 267
DB 131 CTGACTAAATTGGATGGTGACTCCAGGGCGGAGCCGACTAAAGCGTTAAAGAGGCTCT 190

QY 268 GlnLysProIleLysPheValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHis 287
DB 191 GGGAGCCCATTAATTTGTTGGCGTGGGAGCGATTAGAGCCCTTGAGCTTTCTAC 250

QY 288 ProGluArgMetAlaSerArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLys 307
DB 251 CCTGATCGATCCACAGCAGGTCTCGGAATGGGAGAAGTGGTCAATTGTCGAAAG 310

QY 308 AlaGlnGlnAspValAspGlnGluLysAlaLysAspLeuGluLysLysMetArgGluSer 327
DB 311 ACACAAAGAGTGATGATGACCAAGAAAGCTATAGATTACAGAAAGATCATGAGTGA 370

QY 328 SerPheThrLeuAspPheLeuGlnGlnLeuAspGlnValLysAsnLeuGlyProLeu 347
DB 371 AATTTCGACTTCACAGCTTCTTAACAACTCTCAAAACGTTGGGAAAGTGGTCCAG 430

QY 348 AspAspIleMetLysMetIleProGlyMetAsnLysMetLysGlyLeuAspLysLeuAsn 367
DB 431 AGCGCATATTATGGAATGATCCAGCATGAACAGATAACTCCGCA--CAAAATCGA 487

QY 368 MetSerGluLysGlnIleAspHisIleLysAlaIleIleGlnSerMetThrProAlaGln 387
DB 488 GAAGCTGAGAAAGAGACTTGCATTCGTGGATCAATGATCAATGCCATGCTGTGAGAA 547

QY 388 ArgAsnAsnProAspThrLeuAsnValSerArgLysLysArgIle-----AlaLysGly 405
DB 548 AGGAGAACCCAGACTTACTCGCTCAATCACGTGAGAGAGAGATGAAGTGGCTGAGAG 607

QY 406 SerGlyArgSerLeuGlnValAsnArgLeuMetLysGlnPheAsnAspMetLysLys 425
DB 608 TCTGGAAGAGCTGAACAAAGAGGTGAGTCAATTGTTGCCAGCTTTTCCAAATGCGTGT 667

QY 426 MetMetLysGlnPheThrGlyGlyGlyLysGlyLysLys 438
DB 668 CAGATGACAGAGTTGATGGGTATGGTGCAGAGCAAGAA 706

RESULT 13
AW930742
LOCUS AW930742 533 bp mRNA linear EST 18-MAY-2001
DEFINITION ESTJ35685 tomato fruit mature green, TAMU Lycopersicon esculentum
cDNA clone cLEF40N15 5', mRNA sequence.
ACCESSION AW930742
VERSION AW930742.1 GI:8106143
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 533)
AUTHORS Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Rønning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..533
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF40N15"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the lcm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 171 a 86 c 144 g 131 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 8.86e-47 Length: 533
Score: 513.50 Matches: 97
Percent Similarity: 76.27% Conservative: 38
Best Local Similarity: 54.80% Mismatches: 41
Query Match: 22.58% Indels: 1
DB: 1 Gaps: 1

US-09-943-108a-2 (1-455) x AW930742 (1-533)

QY 108 GlyLeuGlnGlyAlaGlyLysThrThrAlaGlyLysLeuAlaLeuLeuMetArgLys 127
DB 5 GGCTACAGAGGTGTGGGAAGCAACTGTAGTGCAGAGCTGTACAGACCTGCTGATTGAC 61

QY 128 LysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArgProAlaAlaIleAsn 147
DB 62 AAGCAGGGTAAGAGTTGCATCTGCTGCTGGAGAGCGTACAGACCTGCTGATTGAC 121

QY 148 GlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSerGluGlyAspGlnVal 167
DB 122 CAACTTGTATTTTGGGTAAACAGGTGATGATGCTGTTATGTCAGCAGGACAGATGTA 181

QY 168 LysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGluHisLeuAspPhe 187
DB 182 AACCCTGCAGAAATAGCCCGCAAGGATTACAGAGGCCCAAAAGAAATGTAGATGA 241

QY 188 ValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeuMetAsnGlnLeuLys 207
DB 242 GTCAATATGGATACAGCTGGAGAGCTTCAGATAGATAAACTATGATGGATGAATTA 301

QY 208 GlnValLysGlnIleAlaLysProAsnGlnIleMetLeuValValAspSerMetThrGly 227
DB 302 GAGCTGAACAGGGTACTGAACCCACAGAGGTTTACTTGTGATGCAATGACTGCG 361

QY 228 GlnAspAlaValAsnValAlaGluSerPheAspAspGlnLeuAspValThrGlyValThr 247
DB 11 Gaps: 2

```

Db 362 CAGAGAGTCAGCTTTGGTCACACATTCATCAATCTCGAAATGGGAATTAAGTGGTGCATC 421
QY 248 LeuThrLysLeuAspGlyThrArgGlyGlyAlaAlaLeuSerIleArgSerValThr 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 422 TTGACGAGCTAGATGAGATTCAGGGGGGAGCAGCTTAAGTGTCAAGGAGGTATCA 481
QY 268 GlnLysProIleLysPheValMetSerGlnLysLeuAspGlyLeuGlu 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 482 NGAAGCCATCAAGCTCGTAGGAGGGGTCAAGTATGAGGACCTTGAA 532

RESULT 14
B0996389
LOCUS
DEFINITION
  B0996389 677 bp mRNA linear EST 22-AUG-2002
  QGG12L07.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
ACCESSION
  B0996389
VERSION
  B0996389.1 GI:22430785
KEYWORDS
  EST.
SOURCE
  Lactuca sativa.
  Lactuca sativa
  Lactuca sativa
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Asteridae; eusterids II; Asterales; Asteraceae; Lactuceae;
  Lactuca.
REFERENCE
  1 (bases 1 to 677)
  Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
  Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
  ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
  Church,S., Jackson,L. and Bradford,K.
  Lettuce and Sunflower ESTs from the Compositae Genome Project
  http://comgenomics.ucdavis.edu/
  Unpublished (2002)
  Contact: Alexander Kozik [R.W.Michelmore]
  Department of Vegetable Crops, R.W.Michelmore Lab
  University of California at Davis (UCD)
  Asmundson Hall, UCD, Davis, CA 95616, USA
  Tel: 1-(530)-742-1742
  Fax: 1-(530)-752-9659
  Email: akozik@catc.org [michelmore@vegmil.ucdavis.edu]
  belongs to contig QG_CA_Contig5305, see http://cgpdb.ucdavis.edu/
  for details
Plate: QGG12 row: L column: 07.
FEATURES
  source
    Location/Qualifiers
      1..677
        /organism="Lactuca sativa"
        /cultivar="L. serriola"
        /db_xref="taxon:4236"
        /clone="QG12L07"
        /lab_host="E.Coli"
        /note="Vector: pBRCNDSfiAB; The library was constructed
        from 10 different sources of RNA from a single genotype.
        Separate cDNAs were generated using primers that
        incorporated unique 5' and 3' tags to distinguish each
        source of RNA. cDNAs were then pooled, size-fractionated,
        directionally cloned into a custom medium-copy vector and
        transformations made with four size classes to minimize
        size bias. Details of each source of RNA and library
        construction can be obtained at http://cgpdb.ucdavis.edu/
        TAG_LIB-QG_EFGHJ lettuce serriola
        TAG_ISSUE-flowers pre-fertilized
        TAG_SEQ-GCTTGACGGG"
BASE COUNT 213 a 100 c 192 g 172 t
ORIGIN
Alignment Scores:
Pred. No.: 7.17e-46 Length: 677
Score: 506.50 Matches: 100
Percent Similarity: 71.43% Conservative: 45
Best Local Similarity: 49.25% Mismatches: 57
Query Match: 22.27% Indels: 1
DB: 14 Gaps: 1

```

```

US-09-943-108A-2 (1-455) x B0996389 (1-677)
QY 3 PheGluGlyLeuSerGluArgLeuGlnAlaThrMetGlnLysMetArgGlyLysGlyLys 22
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 72 TTTGGTCAGGTACACCAAGTGGCTTGAATCCGCTGGACAAGCTTAAGGGGGAAGAGGTC 131
QY 23 LeuThrGluAlaAspIleLysLysIleMetMetArgGluValArgLeuAlaLeuPheGluAla 42
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 132 TTACGAGAGAGAAATATCTGGACCAATAGGAGCATTAAGACGAGCTCTTCTGGAAGCA 191
QY 43 AspValAsnPheLysValValLysGluPheIleLysThrValSerGluArgAlaLeuGly 62
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 192 GATGTTAGCCTTCTCTTGAAGAAGATTCGTTCAAACTGTAAGTGAACAAGCTGTGGT 251
QY 63 SerAspValMetGlnSerLeuThrProGlyGlnGlnValIleLysLeuValGlnAspGlu 82
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 GTTGGTGTGACTAGAGGAGTCAACACGATCAACAATGGTTAAAATTTGTAAGTGATGAA 311
QY 83 LeuThrLysLeuMetGlyGlyGluAsnThrSerIleAsnMetSerAsnLysProProThr 102
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 CTCCTGAGCTAATGGGAGAGAGGTTTCAGCTTAACCTTTGCAAAATCTGGCCCACT 371
QY 103 ValValMetMetValGlyLeuGlnGlyAlaGlyLysThrThrThrAlaGlyLysLeuAla 122
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 372 GTTATCTTACTGCTGCTACAGAGGTGTTGGGAAGACAACCTGTTAGTGCAAAAATTAGCT 431
QY 123 LeuLeuMetArgLysLysTyrAsnLysLysProMetLeuValAlaAlaAspIleTyrArg 142
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 TTGTACCTC---AGAAAACAGGGGAGAGTTCGATGCTGATTCGCTGGAGATGTATAGA 498
QY 143 ProAlaIleAsnGlnLeuGlnThrValGlyLysGlnIleAspIleProValTyrSer 162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 489 CCAGTCGCGATTGATCAACTTCTTATTTGGTAAACAGAGTGATGCTTCTCTGTATGCA 548
QY 163 GluGlyAspGlnValLysProGlnGlnIleValThrAsnAlaLeuLysHisAlaLysGlu 182
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GTAGGACTGATGTAACACCCAGCAATAAGTACAGCAAGGCTTCAAGAGGCTAAAAAG 608
QY 183 GluHisLeuAspPheValIleIleAspThrAlaGlyArgLeuHisIleAspGluAlaLeu 202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 609 AAGAAATGTAGATGTGTTATTAATGATACAGCAGGAGAGCTTCAGATAGACAAAATATG 668
QY 203 MetAsnGlu 205
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 ATGGATGAA 677

RESULT 15
B0996389
LOCUS
DEFINITION
  B0996389 514 bp mRNA linear EST 04-JAN-2002
  PIC1.19_C04.bl_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
ACCESSION
  B0996389
VERSION
  B0996389.1 GI:18061215
KEYWORDS
  EST.
SOURCE
  sorghum.
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
    clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
  1 (bases 1 to 514)
  Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R.,
  Sudman,M. and Pratt,L.H.
  An EST database from Sorghum: plants infected with a compatible
  pathogen
  Unpublished (2002)
  Contact: Cordonnier-Pratt MM
  Laboratory for Genomics and Bioinformatics
  The University of Georgia, Department of Plant Biology
  Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
  Tel: 706 542 1860
  Fax: 706 583 0210
  Email: mmpratt@uga.edu
  Sequences have been trimmed to exclude PolyA, vector, and regions

```

below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 454 POLYA-No.

FEATURES source

Location/Qualifiers
1. .514
/organism="Sorghum bicolor"
/cultivar="BTX623"
/db_xref="taxon:4558"
/clone_lib="Pathogen-infected compatible 1 (PIC1)"
/tissue_type="Leaves"
/dev_stage="4-week-old seedlings infected with Colletotrichum graminicola"
/note="Vector: pBluescript II SK(-) from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old FR421, a sorghum isolate of the anthracnose pathogen Colletotrichum graminicola. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 159 a 91 c 137 g 127 t
ORIGIN

Alignment Scores:

Pred. No.: 1.22e-43 Length: 514
Score: 485.00 Matches: 90
Percent Similarity: 77.25% Conservative: 39
Best Local Similarity: 53.89% Mismatches: 38
Query Match: 21.33% Indels: 0
DB: 13 Gaps: 0

US-09-943-108A-2 (1-455) x BM323434 (1-514)

QY 194 GlyArgLeuHisIleAspGluAlaLeuMetAsnGluLeuLysGluValLysGluIleAla 213
Db 10 GGAACACTGCAGATGATAAATCAATGATGATGAATGAAGAAGTGAAGAGCTGTT 69
QY 214 LysProAsnGluIleMetLeuValValAspSerMetThrGlyGlnAspAlaValAsnVal 233
Db 70 AATCTACAGAGTTCGTGCTGTCATGCCATGACTGSCCAAGAGCTGCACACTA 129
QY 234 AlaGluSerPheAspAspGlnLeuAspValThrGlyValThrLeuThrLysLeuAspGly 253
Db 130 GTCACACCTTCAATATTGAGATTGATGCTGCTATATGACTTAATTTGATGTT 189
QY 254 AspThrArgGlyValAlaLeuSerIleArgSerValThrGlnLysProIleLysPhe 273
Db 190 GACTCCAGGGCGGAGCGGACACTAGTGTAAAGAGCTCTCTGGAGGCCCATCAAGTT 249
QY 274 ValGlyMetSerGluLysLeuAspGlyLeuGluLeuPheHisProGluArgMetAlaSer 293
Db 250 GTTGGCGCTGGGAGCAACATTGGAGACCTTGAGCTTTACCCCTGATCGCATGGCACAG 309
QY 294 ArgIleLeuGlyMetGlyAspValLeuSerLeuIleGluLysAlaGlnAspValAsp 313
Db 310 CGAGTATTGGGAATGGAGATGTACTCTCATTTGTTGAAAGACACAGAAGTTATGCGG 369
QY 314 GlnGluLysAlaLysLeuLysLysMetArgGluSerSerPheThrLeuAspAsp 333
Db 370 CANGAGAGGCTGTAGATTACAGAAAAGATCATGAGTCGAAAAATTCGACTTCACGAC 429
QY 334 PheLeuGluGlnLeuAspGlnValLysAsnLeuGlyProLeuAspAspIleMetLysMet 353

Db 430 TTCTTAACAATCTCAAAATTTCCGAAATGGGTTCATGAGCCGCAATTATTGGAATG 489
QY 354 IleProGlyMetAsnLysMet 360
Db 490 ATGCCAGGCATGAACAAGATA 510

Search completed: February 25, 2003, 03:09:37
Job time : 1571 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:13:23 ; Search time 40 seconds
(without alignments)
1515.726 Million cell updates/sec

Title: US-09-943-108a-2

Perfect score: 2274

Sequence: 1 MAPEGLSERLQATQKMSGK.....GKKGRNQMNLMKGNLPF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2267	99.7	455	AAU00910	S. aureus ffh prot
2	2261	99.4	455	22 AAU37473	Staphylococcus aur
3	2235	98.3	455	22 AAU33924	Staphylococcus aur
4	2235	98.3	455	22 AAU36535	Staphylococcus aur
5	1717.5	75.5	450	23 ABA49131	Listeria monocytog
6	1534	67.5	338	23 ABP39950	Staphylococcus epi
7	1418.5	62.4	472	22 AAU33449	Enterococcus faeca
8	1418.5	62.4	472	22 AAU35268	Enterococcus faeca
9	1315	57.8	518	23 ABA54962	Lactococcus lactis
10	1307	57.5	523	22 AAU37661	Streptococcus pneu

11	1306	57.4	466	22 AAU37932	Streptococcus pneu
12	1306	57.4	523	22 AAU01025	CFE 28 protein seq
13	1305	57.4	284	20 AAU00911	S. aureus ffh prot
14	1297	57.0	523	20 AAU99658	Streptococcus pneu
15	1297	57.0	523	23 AAU79163	Streptococcus pneu
16	1279.5	56.3	521	23 ABP26457	Streptococcus poly
17	1273.5	56.0	521	23 ABP26456	Streptococcus poly
18	1168	51.4	547	22 AAG92001	C glutamicum prote
19	1164.5	51.2	433	22 AAG34670	E. coli cellular p
20	1164.5	51.2	453	22 AAG98879	E. coli growth and
21	1150	50.6	457	22 AAU36355	Pseudomonas aerugi
22	1138	50.0	462	22 AAU35372	Haemophilus influe
23	1127.5	49.6	501	22 AAU38496	Salmonella typhi c
24	1014	44.6	531	22 AAU1288	Propionibacterium
25	869	38.2	448	22 AAU37201	Chlamydia trachoma
26	854	37.6	448	22 AAU35969	Helicobacter pylori
27	846	37.2	420	20 AAU81979	Ehrlichia sp. E74.
28	836	36.8	448	20 AAU34715	Chlamydia pneumonia
29	795	35.0	366	18 AAW20643	H. pylori cytoplasm
30	759.5	33.4	443	22 AAB95153	Putative P. abyssal
31	612.5	26.9	504	23 ABB57149	Mouse ischaemic co
32	592.5	26.1	535	17 AAR91310	Fungal signal reco
33	586	25.8	534	17 AAR91309	Fungal signal reco
34	576.5	25.4	508	22 ABB60834	Drosophila melanog
35	560	24.6	479	21 AAG41615	Arabidopsis thalia
36	548.5	24.1	463	21 AAG41616	Arabidopsis thalia
37	535.5	23.5	404	22 AAG74051	Human colon cancer
38	521	22.9	428	21 AAG41617	Arabidopsis thalia
39	432	19.0	497	22 AAU34785	E. coli cellular p
40	429.5	18.9	329	18 AAW11328	Bacillus subtilis
41	429	18.9	328	23 ABB48889	Listeria monocytog
42	428	18.8	204	18 AAW20559	Helicobacter pylori
43	428	18.8	204	18 AAW24688	H. pylori cytoplasm
44	428	18.8	332	22 AAB96695	Putative P. abyssal
45	426.5	18.8	415	23 ABP39972	Staphylococcus epi

ALIGNMENTS

RESULT 1

AAU00910
ID AAU00910 standard; Protein; 455 AA.

XX AC AAU00910;

XX XX 28-MAY-1999 (first entry)

XX DE S. aureus ffh protein sequence.

XX KW Ffh gene; signal recognition particle; SRP; antimicrobial agent; vaccine;
XX KW Immunological response; gene therapy; infection; otitis media;
XX KW conjunctivitis; toxic shock syndrome; septic arthritis.

XX OS Staphylococcus aureus.

XX PN EP902087-A2.

XX PD 17-MAR-1999.

XX PF 24-AUG-1998; 98EP-0306741.

XX PR 10-SEP-1997; 97US-0927216.

XX PR (SMIK) SMITHKLINE BEECHAM.

XX PI Wallis NG;

XX DR WPI; 1999-169238/15.

XX DR N-PSDB; AAX27221.

XX PT New Staphylococcus aureus Signal Recognition Particle (SRP) with
XX PT protein (ffh) and RNA (ffs) components - the SRP gene and protein

PT useful as diagnostic reagents and for prevention and treatment of
PT Staphylococci infections which cause otitis media, septic arthritis
PT and toxic shock syndrome
XX Claim 11; Page 29-30; 35pp; English.
XX This sequence is the Staphylococcus aureus signal recognition
CC particle (SRP) ffh component. Ffh polynucleotides are useful for
CC diagnosing a disease related to expression of ffh polypeptides by
CC analysing for the presence/amount of ffh protein in a sample due to
CC infection of a micro-organism with the gene, or determining the nucleic
CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful
CC for treatment of an individual in need (polypeptide) of, or needing to
CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and
CC polynucleotides are useful for identifying agonists and antagonists by
CC binding and observing the affect of ffh polypeptide activity, which are
CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments
CC are also useful for inducing an immunological response (T cell/antibody)
CC to protect against disease, by direct administration (vaccine), or via a
CC vector (gene therapy). Anti-fhh antibodies are useful as antagonists, and
CC for protecting against disease. Diseases diagnosed, prevented and treated
CC include those caused by infection, especially bacterial infection,
CC including otitis media, conjunctivitis, toxic shock syndrome, wound
CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and
CC are useful for bathing wounds and implants prior to surgical
CC implantation.
XX
SQ Sequence 455 AA;

Query Match 99.7%; Score 2267; DB 20; Length 455;
Best Local Similarity 99.8%; Pred. No. 2.5e-174;
Matches 454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
Db 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSKPPPTVMVMVGLQAGKTTAGK 120
Db 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSKPPPTVMVMVGLQAGKTTAGK 120
QY 121 LALLMRKKYKPKMLVAADIVPAAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180
Db 121 LALLMRKKYKPKMLVAADIVPAAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKKEIARPNELMIVDSMTGQDQAVNVAESFDDQ 240
Db 181 KEEHLDFVIIDTAGRLHIDEALMNLKEVKKEIARPNELMIVDSMTGQDQAVNVAESFDDQ 240
QY 241 LDVTGVTLTKLDGTRGGAALSIRSVTQKPKFVGNSEKLDGLFHPERMASRIILMGD 300
Db 241 LDVTGVTLTKLDGTRGGAALSIRSVTQKPKFVGNSEKLDGLFHPERMASRIILMGD 300
QY 301 VLSLEKAAQDVQEKAKDLKEMRESSTLDDLEQDQVKNLGLDDIMKMPGNKM 360
Db 301 VLSLEKAAQDVQEKAKDLKEMRESSTLDDLEQDQVKNLGLDDIMKMPGNKM 360
QY 361 KGLDLKNNSEKQIDHAIKIIQSMTPAERNPDITLNVSRKKRIAGSGRSLOEVRNLMKQF 420
Db 361 KGLDLKNNSEKQIDHAIKIIQSMTPAERNPDITLNVSRKKRIAGSGRSLOEVRNLMKQF 420
QY 421 NDKKKMAKQFTGGGKGGKGRNQMNMLKGNLFP 455
Db 421 NDKKKMAKQFTGGGKGGKGRNQMNMLKGNLFP 455

RESULT 2
AAU37473
ID AAU37473 standard; Protein; 455 AA.
XX
AC AAU37473;
XX
DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #1643.
DE
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253623P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR
DR N-PSDB; AAS55332.
XX
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 13066; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, themselves and the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 AA;

Query Match 99.4%; Score 2261; DB 22; Length 455;
Best Local Similarity 99.3%; Pred. No. 7.6e-174;
Matches 452; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
Db 1 MAFGLSERLOATMOKMRGKGLTEADIKIMREVRALFEADVNFVKVKEFIKTVSERA 60
QY 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSKPPPTVMVMVGLQAGKTTAGK 120
Db 61 LGSQVMSLTPGQGVKIVQDELTKLMGENTSTINMSKPPPTVMVMVGLQAGKTTAGK 120
QY 121 LALLMRKKYKPKMLVAADIVPAAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180
Db 121 LALLMRKKYKPKMLVAADIVPAAINOLQTVGKQIDIPVYSEGQVXPQOIVTNALKHA 180

QY 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPEINMLVDSMTGQDAVNVAESFDDQ 240
 DQ 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPEINMLVDSMTGQDAVNVAESFDDQ 240
 QY 241 LDVTGVTLLKLDGTPRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300
 DQ 241 LDVTGVTLLKLDGTPRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300
 QY 301 VLSLIEKAQQVDQEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360
 DQ 301 VLSLIEKAQQVDQEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360
 QY 361 KGLDKLNMESEKQIDHIKAIIOQSMTPAERNNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQF 420
 DQ 361 KGLDKLNMESEKQIDHIKAIIOQSMTPAERNNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQF 420
 QY 421 NDMMKMMKQFTGGGKGKGRKQNMNMLKGMNLPF 455
 DQ 421 NDMMKMMKQFTGGGKGKGRKQNMNMLKGMNLPF 455

RESULT 3
 ID AAU33924 standard; Protein; 455 AA.
 AC AAU33924;
 XX
 DT 14-FEB-2002 (first entry)
 DE Staphylococcus aureus cellular proliferation protein #200.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Staphylococcus aureus.
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207272P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI: 2001-611495/70.
 DR N-PSDB; AAS51783.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 5420; 511pp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 455 AA;

Query Match 98.3%; Score 2235; DB 22; Length 455;
 Best Local Similarity 98.2%; Pred. No. 9.5e-172;
 Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAPEGLSERLOATQKMRGKGLTEADIKIMREVRALFEADVNFVKVEIKTVSERA 60
 DQ 1 MAPEGLSERLOATQKMRGKGLTEADIKIMREVRALFEADVNFVKVEIKTVSERA 60
 QY 61 LGSVMQSLTFGQGVIKIVODELTKLGGENTSTNMGNKPTVVMVGLQGAGKTTTAKG 120
 DQ 61 LGSVMQSLTFGQGVIKIVODELTKLGGENTSTNMGNKPTVVMVGLQGAGKTTTAKG 120
 QY 121 LALLMRKKYKPKMLVAADIYRPAALNQLQTVGKQIDIPVISEGQVKKQIVTNALKHA 180
 DQ 121 LALLMRKKYKPKMLVAADIYRPAALNQLQTVGKQIDIPVISEGQVKKQIVTNALKHA 180
 QY 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPEINMLVDSMTGQDAVNVAESFDDQ 240
 DQ 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPEINMLVDSMTGQDAVNVAESFDDQ 240
 QY 241 LDVTGVTLLKLDGTPRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300
 DQ 241 LDVTGVTLLKLDGTPRGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASRILGMD 300
 QY 301 VLSLIEKAQQVDQEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360
 DQ 301 VLSLIEKAQQVDQEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKMGMLPF 360
 QY 361 KGLDKLNMESEKQIDHIKAIIOQSMTPAERNNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQF 420
 DQ 361 KGLDKLNMESEKQIDHIKAIIOQSMTPAERNNPDTLNVSRKKRIAKGSGRSLOEVNRLMKQF 420
 QY 421 NDMMKMMKQFTGGGKGKGRKQNMNMLKGMNLPF 455
 DQ 421 NDMMKMMKQFTGGGKGKGRKQNMNMLKGMNLPF 455

RESULT 4
 AAU36535
 ID AAU36535 standard; Protein; 455 AA.
 AC AAU36535;
 XX
 DT 14-FEB-2002 (first entry)
 DE Staphylococcus aureus cellular proliferation protein #705.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Staphylococcus aureus.
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207272P.

```
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
DR WPI: 2001-611495/70.
DR N-PSDB; AAS54394.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12128; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 AA;
Query Match 98.34; Score 2235; DB 22; Length 455;
Best Local Similarity 98.28; Pred. No. 9,5e-172;
Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
Db 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
QY 61 LGSVYMSLTPGQVQIKIVQDELTKLMGGENTSIINMSKPPVTVVAVGLQAGKTTAGK 120
Db 61 LGSVYMSLTPGQVQIKIVQDELTKLMGGENTSIINMSKPPVTVVAVGLQAGKTTAGK 120
QY 121 LALIMRRKYNKPMVAADYIRPAINQLOITVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
Db 121 LALIMRRKYNKPMVAADYIRPAINQLOITVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
QY 191 KEELHFDVIIDTAGRLHIDEALMELKEVEIAKPNELMVVDSMTGQDVAVNAESPDQ 240
Db 191 KEELHFDVIIDTAGRLHIDEALMELKEVEIAKPNELMVVDSMTGQDVAVNAESPDQ 240
QY 241 LDVTGVTLLKLDGTRGGAAISIRSVTKPIKFGVMSKLDGLFLFHPERNASRLMGMD 300
Db 241 LDVTGVTLLKLDGTRGGAAISIRSVTKPIKFGVMSKLDGLFLFHPERNASRLMGMD 300
QY 301 VLSLIEKAQODVDQEKADLEKMRGKGLTESFTLDDFLQDLQVKNLGLPDDIMKPIGNKMK 360
Db 301 VLSLIEKAQODVDQEKADLEKMRGKGLTESFTLDDFLQDLQVKNLGLPDDIMKPIGNKMK 360
QY 361 KGLDKLNMSKQIDHIIKAIQISMTPAERNPNPTLNVSRKKRIAGSGRSLQEVNRLMKQF 420
Db 361 KGLDKLNMSKQIDHIIKAIQISMTPAERNPNPTLNVSRKKRIAGSGRSLQEVNRLMKQF 420
QY 421 NDMMKKMMKQFTGGGKKGKGRNQNMNLKGMNLPF 455
```

```
Db 421 NDMMKKMMKQFTGGGKKGKGRNQNMNLKGMNLPF 455
RESULT 5
ABE49131
ID ABE49131 standard; Protein; 450 AA.
XX
AC ABE49131;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #1835.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
PR 11-APR-2000; 2000FR-0004629.
PA (INSP ) INST PASTEUR.
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Cherouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Gobel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI: 2002-010914/01.
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
XX Claim 6; SEQ ID No 1836; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 450 AA;
Query Match 75.5%; Score 1717.5; DB 23; Length 450;
Best Local Similarity 72.7%; Pred. No. 4.8e-130;
Matches 331; Conservative 64; Mismatches 55; Indels 5; Gaps 1;
QY 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMREVRLLAEADVNFVKVKEFIKTVSERA 60
|||||: ||| ||:||||: |||:| ||||||| |||||||:|||||
```

Db 1 MAFEGLAGRQETMKNKTRGKGVNADYKEMMREVRLLALEADVNFVKVQKFIKTVSERA 60
QY 61 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 120
Db 61 VGDVYKSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 120
QY 121 LALLMRKYNKKPMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQPOQIVTNALKHA 180
Db 121 LALLMRKYNKKPMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQPOQIVTNALKHA 180
QY 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240
Db 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTQKPIKFMGSEKLDGLELPHPERMASRILGMD 300
Db 241 LEITGVTITKLDGTRGGAALSIRSVTQKPIKFMGSEKLDGLELPHPERMASRILGMD 300
QY 301 VLSLIEKAQDVQDEKAKDEKKMRESSFTLDDFLEOLDQVKNLGPLDDIMKKIPGNKM 360
Db 301 VLSLIEKAQDVQDEKAKDEKKMRESSFTLDDFLEOLDQVKNLGPLDDIMKKIPGNKM 360
QY 361 KGLDKLNMSKOIDHIAIKIQTSMTPAERNPDITLVNRRKRIAGSGRSLOEVNRLAKQF 420
Db 361 KGLDNMVDQOLGHEIAIKSMTKNEKNDPDIINASRRKRIAGSGRPVQEIINLLKQF 420
QY 421 NDKMKMKOFYGGGKGKGGKGNOMQNLKGNLFP 455
Db 421 AEMKKMKOMTGGGKGKGGK-----NPFGNFKMPF 450

RESULT 6
ABP39950
ID ABP39950 standard; Protein; 338 AA.
XX AC ABP39950;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4795.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2002-381255/41.
XX DR N-PSDB; ABN92495.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PT polypeptide, useful for diagnosing and treating bacterial infections -
XX FS Disclosure; SEQ ID 4795; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to

CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX SQ Sequence 338 AA;
Query Match 67.5%; Score 1534; DB 23; Length 338;
Best Local Similarity 92.5%; Pred. No. 2e-115;
Matches 307; Conservative 16; Mismatches 9; Indels 0; Gaps 0;
QY 1 WAFGLSERLOATMOKRGKGLTENDIKIMREVRLLALEADVNFVKVQKFIKTVSERA 60
Db 6 WAFGLSDRLQATMQKRGKGVTEADIKTMREVRLLALEADVNFVKVQKFIKTVSERA 65
QY 61 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 120
Db 66 LGSVMQSLTPGOQVIKIVQDELTKLMGGENTSIINMSNKPPTVVMVGLQAGKTTTAGK 125
QY 121 LALLMRKYNKKPMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQPOQIVTNALKHA 180
Db 126 LALLMRKYNKKPMLVAADYRPAAINOLQTVGKQIDIPVYSEGQVQPOQIVTNALKHA 185
QY 181 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 240
Db 186 KEEHLDVFIIDTAGRLHIDEALMNLKEVKEIAKPNEIMLVDSMTGQDANVAESFDDQ 245
QY 241 LDVTGVTITKLDGTRGGAALSIRSVTQKPIKFMGSEKLDGLELPHPERMASRILGMD 300
Db 246 LDVSGVTITKLDGTRGGAALSIRSVTQKPIKFMGSEKLDGLELPHPERMASRILGMD 305
QY 301 VLSLIEKAQDVQDEKAKDEKKMRESSFTLD 332
Db 306 VLSLIEKAQDVQDEKAKDEKKMRESSFTLD 337

RESULT 7
AAU33449
ID AAU33449 standard; Protein; 472 AA.
XX AC AAU33449;
XX DT 14-FEB-2002 (first entry)
XX DE Enterococcus faecalis cellular proliferation protein #85.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Enterococcus faecalis.
XX PN WC200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2001; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207272P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS51308.

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 4945; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 AA;

Query Match 62.4%; Score 1418.5; DB 22; Length 472;
Best Local Similarity 59.0%; Pred. No. 6.5e-106;
Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAFEGSLERLOATMOKMRGKGLTADIKIMREVELALFEADVNFKVVEFIKTVSERA 60
DB 1 MAFESLIRLQOAMSKIRKRGKGVADVKEMREIRLALLEADVNLQVVKDFTRVRERA 60
QY 61 LGSVMQSLTPGQGVKIVQDELTILMGENTSSINMSKPPVTVMVGLQAGKTTAGK 120
DB 61 VGVVELESLSPAQGVKIVDELTILGSETVELNKSPIPTVINMTGLQAGKTTFTCK 120
QY 121 LALLMKRYKKNKPMVAADIVRPAINQLOTGVKQIDIPVYSESGQVKKPOQIVTNALKHA 180
DB 121 LAKHLMKTENARPLLIAGDVYRPAADQLKVLQGLEVPFDMGTDPANPEIVRQGLALA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNELKEVKEIAKPNEILVVDVSDMTGDAVNVAESFDDQ 240
DB 181 KEEKNDYVILDTAGRLHIDEALMDELKOLKELANPEILLVVDVMTGQDAVNVAESFNEQ 240
QY 241 LDVTGVTKLDGTRGGAALSTRVQKPKFVGMSEKLDGLELPHPERVAKSILQMGD 300
DB 241 LGITGVWTKMDGTRGGAALSTRVQKPKFVGSGEKLTDLFPHPRMSRILQMGD 300
QY 301 VLSLIEKAQQVDQERAKLEKMRSEFTLDDFLEQLDDVKNLGLPDDIMKMPGNKM 360
DB 301 MFLIEKAQQYDEKKAELAAQKMNSEFDFNDFIEQLDQVGMGPIEDLLKMTPGMSNM 360
QY 361 KGLDKLNMSEKQIDHITKAIQSTPTFAERNPTLNVSRKKRIAGSGRSLOEVRNLMKQF 420
DB 361 PGIENVKVDPRDVARKEAWVLSMTPAERENPDLLNPSRRRIIAAGSGNSVVEVNRMIKQF 420
QY 421 NDMMKMKQFT-----GGGKGGKGNQMNK 449
DB 421 KESKMMQOMSKGDMNIPGMDQLGGVKGKLGKM-AMNPMK 462

RESULT 8
AAU35268
ID AAU35268 standard; Protein; 477 AA.
AC
XX AAU35268;
XX
QT 14-FEB-2002 (first entry)
XX

DE Enterococcus faecalis cellular proliferation protein #555.
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Enterococcus faecalis.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
PA
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX WPI; 2001-611495/70.
DR N-PSDB; AAS53127.
DR
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10861; 511pp; English.
PS
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 477 AA;

Query Match 62.4%; Score 1418.5; DB 22; Length 477;
Best Local Similarity 59.0%; Pred. No. 6.6e-106;
Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAFEGSLERLOATMOKMRGKGLTADIKIMREVELALFEADVNFKVVEFIKTVSERA 60
DB 1 MAFESLIRLQOAMSKIRKRGKGVADVKEMREIRLALLEADVNLQVVKDFTRVRERA 60
QY 61 LGSVMQSLTPGQGVKIVQDELTILMGENTSSINMSKPPVTVMVGLQAGKTTAGK 120
DB 61 VGVVELESLSPAQGVKIVDELTILGSETVELNKSPIPTVINMTGLQAGKTTFTCK 120
QY 121 LALLMKRYKKNKPMVAADIVRPAINQLOTGVKQIDIPVYSESGQVKKPOQIVTNALKHA 180
DB 121 LAKHLMKTENARPLLIAGDVYRPAADQLKVLQGLEVPFDMGTDPANPEIVRQGLALA 180
QY 181 KEEHLDFVIIDTAGRLHIDEALMNELKEVKEIAKPNEILVVDVSDMTGDAVNVAESFDDQ 240

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 523 AA;

Query Match 57.5%; Score 1307; DB 22; Length 523;

Best Local Similarity 55.1%; Pred. No. 7.3e-97;

Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSRLQATMQRKRGKGLTEADIKIMRVRRLALFEADVNFVKVEFIKTVSERA 60
 DB 1 MAFESLTERLQWFKNLKRGKGLTESDQVQATKEIRLALFEADVLPVVKDFIKVVERA 60
 QY 61 LGSVYMSLTGPGQVQIKIVDELTKLKGENTSNMKNPPTVVMVGLQGAGKTTTAGK 120
 DB 61 VGHEVIDTLNPAQOIIVDELTAVLGSDAEIKSPKTIPTIMVGLQGAGKTTTAGK 120
 QY 121 LALLMRKYNKPKMLVAADIRPAALNQLQVTKGIDIPVYSGDQVKKPQIVTNALKHA 180
 DB 121 LANKLKKEENARPLMVAADIRPAALDQLKLGQIDVPVGLTGVPAVEIVRQGLEQA 180
 QY 181 KEEHLDFVIIDTAGRLHIDPALMNLKEVKEIKAPNEIMLVDSMTGQDAVNVASFDDQ 240
 DB 181 QTNHNDYVLIDTAGRLQIDELLNLRDVKALQAQNEILLVDMITGQDAVNVAREFNAQ 240
 QY 241 LDVTGVTITKLDGTRGAALSRVSTQPKIFVGMSEKLDGLELPHPERMASRLTGMGD 300
 DB 241 LEVTGVILTKLDGTRGAALSVRHTGPKIFGTGKEITDIEITHPDRMSRLTGMGD 300
 QY 301 VLSLEKAAQDVQDEKAKLEKMRSSFTLDDFLEQLDQVKNLGPDLDDIMKMPGNMKM 360
 DB 301 MLTLEKASQYDEKALENAEKMRNTDFDNFDLDQVQNGPMDLAKMIFGMANN 360
 QY 361 KGLDKLNSEKQIDHIKAIQSWTPAERNNDPTLVNSKKRIKAGRSLSQEVNRLMKQF 420
 DB 361 PALQNMKYDERQIARKRAIVSSPTPERENPDLLNPSRRRIAGSGNTFVEVNFKIDF 420
 QY 421 NDMKKMKAQFTGCGKGGKGNQMNKMG-----NLP 454
 DB 421 NQAKOLMGVMSG-----DANKMKMGQMGINDNLP 450

RESULT 11

ID AU037932

XX AAU037932 standard; Protein; 466 AA.

AC AAU037932;

XX AAU037932;

DT 14-FEB-2002 (first entry)

XX Streptococcus pneumoniae cellular proliferation protein #361.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Streptococcus pneumoniae.

XX WO200170955-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 21-MAR-2001; 2001WO-US09180.
 PF
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITFA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 PI Yamamoto RT, Xu HH;
 PI
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS55791.
 DR
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX Example 3; Seq ID No 13525; 511pp; English.
 PS
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 466 AA;

Query Match 57.4%; Score 1305; DB 22; Length 466;

Best Local Similarity 54.9%; Pred. No. 7.5e-97;

Matches 252; Conservative 91; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSRLQATMQRKRGKGLTEADIKIMRVRRLALFEADVNFVKVEFIKTVSERA 60
 DB 1 MAFESLTERLQWFKNLKRGKGLTESDQVQATKEIRLALFEADVLPVVKDFIKVVERA 60
 QY 61 LGSVYMSLTGPGQVQIKIVDELTKLKGENTSNMKNPPTVVMVGLQGAGKTTTAGK 120
 DB 61 VGHEVIDTLNPAQOIIVDELTAVLGSDAEIKSPKTIPTIMVGLQGAGKTTTAGK 120
 QY 121 LALLMRKYNKPKMLVAADIRPAALNQLQVTKGIDIPVYSGDQVKKPQIVTNALKHA 180
 DB 121 LANKLKKEENARPLMVAADIRPAALDQLKLGQIDVPVGLTGVPAVEIVRQGLEQA 180
 QY 181 KEEHLDFVIIDTAGRLHIDPALMNLKEVKEIKAPNEIMLVDSMTGQDAVNVASFDDQ 240
 DB 181 QTNHNDYVLIDTAGRLQIDELLNLRDVKALQAQNEILLVDMITGQDAVNVAREFNAQ 240
 QY 241 LDVTGVTITKLDGTRGAALSRVSTQPKIFVGMSEKLDGLELPHPERMASRLTGMGD 300
 DB 241 LEVTGVILTKLDGTRGAALSVRHTGPKIFGTGKEITDIEITHPDRMSRLTGMGD 300

XX XX

PT protein (ffh) and RNA (ffs) components - the SRP gene and protein useful as diagnostic reagents and for prevention and treatment of PT

SRP
tion

PT Staphylococci infections which cause otitis media, septic arthritis
 PT and toxic shock syndrome
 XX
 PS Claim 11; Page 31-32; 35pp; English.
 XX
 CC This sequence is the Staphylococcus aureus signal recognition
 CC particle (SRP) ffh component. Ffh polynucleotides are useful for
 CC diagnosing a disease related to expression of ffh polypeptides by
 CC analysing for the presence/amount of ffh protein in a sample due to
 CC infection of a micro-organism with the gene, or determining the nucleic
 CC acid sequence encoding ffh. Ffh polypeptides and antagonists are useful
 CC for treatment of an individual in need (polypeptide) of, or needing to
 CC inhibit (antagonist) ffh polypeptide levels. Ffh polypeptides and
 CC polynucleotides are useful for identifying agonists and antagonists by
 CC binding and observing the affect of ffh polypeptide activity, which are
 CC potential anti-microbial agents. Ffh polypeptides and antigenic fragments
 CC are also useful for inducing an immunological response (T cell/antibody)
 CC to protect against disease, by direct administration (vaccine), or via a
 CC vector (gene therapy). Anti-fhh antibodies are useful as antagonists, and
 CC for protecting against disease. Diseases diagnosed, prevented and treated
 CC include those caused by infection, especially bacterial infection,
 CC including otitis media, conjunctivitis, toxic shock syndrome, wound
 CC infection and septic arthritis. Ffh polypeptides are antimicrobial, and
 CC are useful for bathing wounds and implants prior to surgical
 CC implantation.
 XX

Sequence 264 AA;

Query Match 57.4%; Score 1305; DB 20; Length 264;
 Best Local Similarity 100.0%; Pred. No. 4.1e-97;
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60
 DB 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60
 QY 61 LGSVDWQSLTPGQGVKIQVDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120
 DB 61 LGSVDWQSLTPGQGVKIQVDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120
 QY 121 LALLMRKKYKPKMLVAADIYRPAALNQLTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180
 DB 121 LALLMRKKYKPKMLVAADIYRPAALNQLTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180
 QY 181 KEHLDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEMLVYDSMTGQDAVNVAFSFDQ 240
 DB 181 KEHLDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEMLVYDSMTGQDAVNVAFSFDQ 240
 QY 241 LDVGTGTLKLDGDRGGGALSIR 264
 DB 241 LDVGTGTLKLDGDRGGGALSIR 264

RESULT 14

AAW99658
 ID AAW99658 standard; Protein; 523 AA.

XX
 AC
 XX
 XX
 DT
 XX
 XX
 DE

Streptococcus pneumoniae ffh protein.

XX Streptococcus pneumoniae; ffh; fifty-four homologue; antibacterial;
 KW infection; otitis media; conjunctivitis; bacteraemia; sinusitis;
 KW pleural empyema; endocarditis; meningitis.

XX Streptococcus pneumoniae.

XX
 PN EP900843-A2.
 XX
 XX 10-MAR-1999.
 PD
 XX

PF 20-AUG-1998; 98EP-0306685.
 XX
 PR 02-SEP-1997; 97US-0923772.
 XX
 PA (SMIX) SMITHKLINE BEECHAM.
 XX
 XX Black WT;
 PI
 XX
 XX WPI; 1999-155936/14.
 DR N-PSDB; AAX19484.
 XX
 PT New Streptococcus pneumoniae Fifty-Four Homologue (Ffh) polypeptide
 PT and polynucleotide - useful as diagnostic reagents and for
 PT prevention and treatment of Streptococci infections, which cause
 PT otitis media, sinusitis and conjunctivitis
 XX
 PS Claim 12; Page 6; 21pp; English.

XX The present sequence represents the Streptococcus pneumoniae fifty-four
 CC homologue (ffh) protein, which is a component of the protein secretory
 CC apparatus in bacteria, and the bacterial homologue of the eukaryotic
 CC Signal Recognition Particle. Ffh proteins and polynucleotides are useful
 CC for diagnosing diseases related to over or underexpression of ffh protein
 CC by identifying mutations in the ffh gene, or determining ffh protein or
 CC mRNA expression levels due to an infection of an organism with the ffh
 CC gene. They can diagnose the stage and type of infection. Ffh proteins are
 CC also useful for screening for compounds which affect activity of the
 CC protein by measuring the binding to ffh protein and observing the
 CC stimulation or inhibition of the protein function. These can be used in
 CC treatment to inhibit (antagonist i.e. antibacterial drugs) or enhance
 CC (agonist) ffh activity, in addition to direct administration of ffh
 CC proteins to treat conditions associated with a lack of ffh protein, or
 CC direct administration of antisense sequences to prevent expression. Ffh
 CC proteins (administered directly, in a vector and as a vaccine) and
 CC antibodies induce an immune response to immunise and prevent disease.
 CC Diseases diagnosed, prevented or treated include: bacterial infections,
 CC especially Streptococcus pneumoniae infections, which cause otitis media,
 CC conjunctivitis, bacteraemia, sinusitis, pleural empyema, endocarditis and
 CC especially meningitis. Ffh proteins, polynucleotides and their
 CC (antagonists) can prevent adhesion of bacteria to matrix proteins, and
 CC are useful for use on wounds and body implants to prevent bacterial
 CC infection.
 XX

Sequence 523 AA;

Query Match 57.0%; Score 1297; DB 20; Length 523;
 Best Local Similarity 54.7%; Pred. No. 4.7e-96;
 Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMQKRGKGLTEADIKIMREVRALFEADYNFKVKEFIKTVSERA 60

DB 1 MAFESTLERLQNVFKNLKKGKISDSVQEAETKEIRLALLADYALFVVKDFIKKVRERA 60

QY 61 LGSVDWQSLTPGQGVKIQVDELTKLMGGENTSINNSKNKPTVVMVGLQGAGKTTAGK 120

DB 61 VGEHVIDTLNPAQQLIKLVEELTAVLGSDTAEIILKSPKPTIIMVGLQGAGKTTAGK 120

QY 121 LALLMRKKYKPKMLVAADIYRPAALNQLTVGKQIDIPVYSEGQVKKPQOIVTNALKHA 180

DB 121 LANKKKEENARPLMAADIYRPAALDQLKLGQQLIDVPEFALGTEVPAVEIVROGLQA 180

QY 181 KEHLDFVIIDTAGRLHIDEALMNEELKEVKEIAKPNEMLVYDSMTGQDAVNVAFSFDQ 240

DB 181 OTVHNDVILDTAGRLQIDELMNEELRVKLAQPNELLVVDAMTQGAANVAREFNAQ 240

QY 241 LDVGTGTLKLDGDRGGGALSIRVTKPIKFGVMSKLDGLFELFHPERMASRLIGMD 300

DB 241 LEVTGVILKIDGDRGGAALSVRHITGPIFTGTGKIDTIDTFHPDMSRLIGMD 300

QY 301 VLSLIEKAQQVDQEKAKDLKXKRESFTLDDFLEQLDQVKNLGLDDIMKIPGMNKM 360

DB 301 MLTIEKASQVEDEQKALENAEKRENTFFDNFDQLDQVQNNGMEDLLKWIPIGMANN 360

Search completed: February 21, 2003, 15:17:57
Job time : 44 secs

Search completed: February 21, 2003, 15:17:57
Job time : 44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:17:12 ; Search time 18 seconds
(without alignments)
743.746 Million cell updates/sec

Title: US-09-943-108a-2

Perfect score: 2274
Sequence: 1 MAFEGLSERLOATMOKMRGK.....GKKGRNQMNLMGNLPPF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2274	100.0	455	4	US-09-035-382-2
2	1534	67.5	338	4	US-09-134-001C-4795
3	1305	57.4	264	4	US-09-035-382-4
4	1297	57.0	523	2	US-08-923-772-2
5	1297	57.0	523	4	US-09-385-287-2
6	845	37.2	420	4	US-09-056-047-8
7	586	25.8	534	2	US-08-317-401E-2
8	578	25.4	552	2	US-08-317-401E-4
9	429.5	18.9	329	4	US-08-981-527A-8
10	426.5	18.8	415	4	US-09-134-001C-4817
11	421	18.5	416	4	US-09-007-476-2
12	397.5	17.5	425	2	US-08-986-963-2
13	397.5	17.5	430	3	US-09-007-484-2
14	397.5	17.5	430	4	US-09-309-682-2
15	390	17.2	277	3	US-09-007-484-4
16	390	17.2	277	4	US-09-309-682-4
17	204.5	9.0	105	4	US-08-981-527A-13
18	179.5	7.9	105	4	US-08-981-527A-12
19	149.5	6.6	129	4	US-08-981-527A-11
20	136	6.0	606	4	US-08-477-831C-2
21	135.5	6.0	631	4	US-08-477-831C-11
22	132	5.8	1010	4	US-09-134-001C-5178
23	131	5.8	10182	4	US-09-134-001C-3159
24	130.5	5.7	1786	4	US-08-973-462-8
25	130	5.7	3696	4	US-09-134-001C-5080
26	127.5	5.6	1093	4	US-09-315-793-52
27	124	5.5	878	4	US-09-134-001C-4378

28	123	5.4	118	4	US-08-981-527A-10	Sequence 10, Appl
29	122.5	5.4	800	3	US-08-776-265-3	Sequence 3, Appl
30	119.5	5.3	423	2	US-08-290-731C-10	Sequence 10, Appl
31	119.5	5.3	1319	2	US-08-290-731C-2	Sequence 2, Appl
32	119.5	5.3	1336	2	US-08-290-731C-6	Sequence 6, Appl
33	119	5.2	593	2	US-08-591-079-8	Sequence 8, Appl
34	118.5	5.2	512	4	US-09-134-001C-4349	Sequence 4349, Ap
35	118.5	5.2	728	4	US-09-134-001C-4968	Sequence 4968, Ap
36	118	5.2	2482	1	US-08-328-254-6	Sequence 6, Appl
37	118	5.2	3248	1	US-08-353-700-1	Sequence 1, Appl
38	118	5.2	3248	5	PCT-US95-16216-1	Sequence 1, Appl
39	117	5.1	480	2	US-08-913-477-17	Sequence 17, Appl
40	116.5	5.1	573	4	US-08-235-836C-112	Sequence 112, App
41	116	5.1	2285	4	US-09-308-375-2	Sequence 2, Appl
42	115.5	5.1	454	4	US-09-134-001C-3547	Sequence 3547, Ap
43	115.5	5.1	804	4	US-09-134-001C-5218	Sequence 5218, Ap
44	115.5	5.1	1886	4	US-08-938-105-3	Sequence 3, Appl
45	115.5	5.1	1939	4	US-09-310-187A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-035-382-2

; Sequence 2, Application US/09035382

; Patent No. 6284515

; GENERAL INFORMATION:

; APPLICANT: Black, Michael T.

; TITLE OF INVENTION: SIGNAL RECOGNITION PARTICLE POLYPEPTIDES

; TITLE OF INVENTION: AND POLYNUCLEOTIDES

; FILE REFERENCE: GM50035

; CURRENT APPLICATION NUMBER: US/09/035,382

; CURRENT FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: 60/057,890

; EARLIER FILING DATE: 1997-09-03

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 455

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-035-382-2

Query Match Similarity 100.0%; Score 2274; DB 4; Length 455;

Best Local Similarity 100.0%; Pred. No. 1.2e-190;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMRREVRLALFEADVNFVKVKEFIKTVSERA 60

Db 1 MAFEGLSERLOATMOKMRGKGLTEADIKIMRREVRLALFEADVNFVKVKEFIKTVSERA 60

Qy 61 LGSVDYMQSLTPGQGVTKIVQDELTKLMGENTSIINMSNKPTVTVMVGLQGAGKTTAGK 120

Db 61 LGSVDYMQSLTPGQGVTKIVQDELTKLMGENTSIINMSNKPTVTVMVGLQGAGKTTAGK 120

Qy 121 LALLMRKYNKPKMLVAADITPPAALNQLQVGVKIDIPVYSEGQVQKPOQIVTNALKHA 180

Db 121 LALLMRKYNKPKMLVAADITPPAALNQLQVGVKIDIPVYSEGQVQKPOQIVTNALKHA 180

Qy 181 KEEHLDFVIIDTAGRLHIDPALMNLKEVKEIAKPEINMLVDSMTGDAVNAESFDDQ 240

Db 181 KEEHLDFVIIDTAGRLHIDPALMNLKEVKEIAKPEINMLVDSMTGDAVNAESFDDQ 240

Qy 241 LDVTGVTITKLDGTRGGAALSIRSVTKPKIKFVGMSEKLDGLELPHFERWASRLGMGD 300

Db 241 LDVTGVTITKLDGTRGGAALSIRSVTKPKIKFVGMSEKLDGLELPHFERWASRLGMGD 300

Qy 301 VLSLIEKAQQDQDQEKAKLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMTIPGNMK 360

Db 301 VLSLIEKAQQDQDQEKAKLEKKMRESSFTLDDFLEQLDQVKNLGPLDDIMKMTIPGNMK 360

Qy 361 KGLDKLNMSEKQIDHIIKAIIOQSWPAERNPDITLNVSKRKRIAKSGRSLQEVRLMKQF 420

Db 361 KGLDKLNMSEKQIDHIIKAIIOQSWPAERNPDITLNVSKRKRIAKSGRSLQEVRLMKQF 420

US-08-923-772-2

```
Query Match          57.0%; Score 1297; DB 2; Length 523;
Best Local Similarity 54.7%; Pred. No. 3.4e-105;
Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMQRKMGKGLTEADIKIMREVRRLALFEADYNFVVKVEFKTVSERA 60
DB 1 MAFESUTERLQNFVKNLRKKKISSESDVQEAATKEIRLALLLEADVALPVVKDFIKKVERA 60

QY 61 LGSVDQSLTPGQQVYIKIVQDELTKMGENTSNMKNRPVVMVGLQAGKTTAGK 120
DB 61 VGEHVIDTLNPAQOIIVDELTAVLGSDTAELIKSPKTIIMMVGLOGAGKTTAGK 120

QY 121 LALLMRKKYKPKMLVAADIYRPAAINQLOTGWKQIDIPVYSEGQVKKPOIVTNALKHA 180
DB 121 LANKLKEENARP:MTAAADLYRPAIDQLKTLGQIDVPVFPALGTEVPAVEIVRQGLEQA 180

QY 181 KEHLDFVIDTAGRLHIDEALMNEKVEKTAIKNELVVDVMTGODAVNVAESFDDQ 240
DB 181 QTNHNDYVLDITAGRLQIDELLMNELRDVKVLAQNEILLVVDAMIGQEAANVAREFNAQ 240

QY 241 LDVGTGVLTKDGTGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASSRIILGMD 300
DB 241 LEVTVGVLTKDGTGGAALSIRSVTKPIKFGTGTGKIDTLETFHEDRMSSRIILGMD 300

QY 301 VLSLIEKAQDDVQDEKADLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKIPGMNM 360
DB 301 MLTLIEKASQYDEQKALEMAEKRENTDFNDFIDQLDQVQNGMPMEDLLKIPGMANN 360

QY 361 KGLDKLNSEKQIDHIAKIQSTWPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKOF 420
DB 361 PALQNMKVDERQIAKRAIVSSHTSERENPDLNPSRRRIAGSGNTFVEVNFKIDF 420

QY 421 NDMMKMKOFTGGGKGGKGRNQNMQLKGM-----NLP 454
DB 421 NOAKQLMOGVMSG-----DMNKKMKQMGINPNLNP 450

RESULT 5
US-09-385-287-2
; Sequence 2, Application US/09385287
; Patent No. 6350857
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: NOVEL fff
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,287
; FILING DATE: 30-Aug-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/923,772
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
```

```
TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-385-287-2

Query Match          57.0%; Score 1297; DB 4; Length 523;
Best Local Similarity 54.7%; Pred. No. 3.4e-105;
Matches 251; Conservative 91; Mismatches 103; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMQRKMGKGLTEADIKIMREVRRLALFEADYNFVVKVEFKTVSERA 60
DB 1 MAFESUTERLQNFVKNLRKKKISSESDVQEAATKEIRLALLLEADVALPVVKDFIKKVERA 60

QY 61 LGSVDQSLTPGQQVYIKIVQDELTKMGENTSNMKNRPVVMVGLQAGKTTAGK 120
DB 61 VGEHVIDTLNPAQOIIVDELTAVLGSDTAELIKSPKTIIMMVGLOGAGKTTAGK 120

QY 121 LALLMRKKYKPKMLVAADIYRPAAINQLOTGWKQIDIPVYSEGQVKKPOIVTNALKHA 180
DB 121 LANKLKEENARP:MTAAADLYRPAIDQLKTLGQIDVPVFPALGTEVPAVEIVRQGLEQA 180

QY 181 KEHLDFVIDTAGRLHIDEALMNEKVEKTAIKNELVVDVMTGODAVNVAESFDDQ 240
DB 181 QTNHNDYVLDITAGRLQIDELLMNELRDVKVLAQNEILLVVDAMIGQEAANVAREFNAQ 240

QY 241 LDVGTGVLTKDGTGGAALSIRSVTKPIKFGVMSKLDGLFHPERMASSRIILGMD 300
DB 241 LEVTVGVLTKDGTGGAALSIRSVTKPIKFGTGTGKIDTLETFHEDRMSSRIILGMD 300

QY 301 VLSLIEKAQDDVQDEKADLEKKMRESSFTLDDFLEQLDQVKNLGLPDDIMKIPGMNM 360
DB 301 MLTLIEKASQYDEQKALEMAEKRENTDFNDFIDQLDQVQNGMPMEDLLKIPGMANN 360

QY 361 KGLDKLNSEKQIDHIAKIQSTWPAERNPDTLVNSRKKRIAKGSGRSLOEVNRLMKOF 420
DB 361 PALQNMKVDERQIAKRAIVSSHTSERENPDLNPSRRRIAGSGNTFVEVNFKIDF 420

QY 421 NDMMKMKOFTGGGKGGKGRNQNMQLKGM-----NLP 454
DB 421 NOAKQLMOGVMSG-----DMNKKMKQMGINPNLNP 450

RESULT 6
US-09-066-047-8
; Sequence 8, Application US/09066047A
; Patent No. 6306394
; GENERAL INFORMATION:
; APPLICANT: MURPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
; USE OF GRANULOCYTIC ERHLICHIA
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,047A
```

```
/
/ FILING DATE: 24-Apr-1998
/ CLASSIFICATION: <Unknown>
/
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 60/044,869
/ FILING DATE: 25-APR-1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Superko, Colleen
/ REGISTRATION NUMBER: 39,850
/ REFERENCE/DOCKET NUMBER: 106,941.156
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 526-6000
/ TELEFAX: (617) 526-5000
/
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 420 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-066-047-8

Query Match 37.2%; Score 846; DB 4; Length 420;
Best Local Similarity 40.8%; Pred. No. 6.3e-66;
Matches 173; Conservative 98; Mismatches 143; Indels 10; Gaps 3;

QY 3 FEGLSERLQATQMKRGKGLFEADIKIMREVRALFEADVNFVKYKFIKTVSERALG 62
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 2 FNSLTKGSSALQSLKREISKDFDLVIDTQALDADVNLGVVDVFENKRSKIVG 61
QY 63 SDVMQSLTPGQQVYIKIVODELTKMGENTSIINMSKPPVAVMVGLOGAGKTTAGKLA 122
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 62 GDVKGVLPEQMWIKRIEELIENLGNESKALDLKGKIPAVIMVVGLOGVGTINTVKVA 121
QY 123 LLMRKYNKKPLVVAADIRPAINQLOTGVKQIDIPVYSEGQVQKQIYVINALKHAKE 182
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 122 LRL-KKDSKNPLVASLDVYPAARQLVLADGVGIDSLIVEEQKPLDIKRAMREARL 180
QY 183 EHLDSVITDAGRHLHIDEALMNELEKVEIAKPEIMLVVDSMTGODAVNAESFDQLD 242
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 181 KGHVVLLDTAGRLHINQMDLCKVYKESPAEIVLVVDSLMGODATVYKREBELG 240
QY 243 VTGVTLTKLGDTRGGAALSRSVTKPKIKVGVSEKLDGLFLFHPERNASRIILGMDVYL 302
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 241 ITGTFITRADGPRGGAISLMKIVAGCPKFKMSTGKPELDLDFYPDRIARRMLMGSDVA 300
QY 303 SLIEKAQDVQDEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLDDIMKMPGK---- 357
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 301 SLVEKAVAGKDTINELQAKKGFLLDLVQLKALNKGGINIMKFFIPAGNDIK 360
QY 358 NKMKGDLKNNSEKQIDHIKAIQSWTPAERNPDITLVNRSKRRIAGSGRSLSQEVNRLM 417
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 361 RKVAGT---ADDSKVDMTAIINSMTKQERANPEILNGAKARIAGAGVKVDVAVNALL 416
QY 418 KQFN 421
Db 417 KQFN 420

RESULT 7
US-08-317-401E-2
; Sequence 2, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
```

```
/
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10174-6401
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/317,401E
/ FILING DATE: 03-October-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Harrington, James J.
/ REGISTRATION NUMBER: 38,711
/ REFERENCE/DOCKET NUMBER: 4248,000-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212 867 0123
/ TELEFAX: 212 867 0298
/
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 534 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ ORIGINAL SOURCE:
/ ORGANISM: Aspergillus niger
/
/ US-08-317-401E-2
/
/ Query Match 25.8%; Score 586; DB 2; Length 534;
/ Best Local Similarity 33.6%; Pred. No. 4.6e-43;
/ Matches 155; Conservative 75; Mismatches 213; Indels 18; Gaps 9;
/
/ QY 1 MAPGSLERLQATQMKRGKGLFEAD- IKIMREVRALFEADVNFVKYKFIKTVSER 59
/ Db 1 MVLQDLGRRINAAVNDLRTSNLDEKQAFDDMIKETAALLSADVNVNLSVQSRKSIKSS 60
/ QY 60 ALGSDVMQSLTPGQQVYIKIVODELTKMGENTSIINMSKPPVAVMVGLOGAGKTTAG 119
/ Db 61 VNFASLPPAVNKKRLQKAVFDELVSLVDPHAFPRPKKGRSNVIMFVGLQGAGKTITCT 120
/ QY 120 KLALLMKKYNK---KPLMVAADIYRPAINQLOTGVKQIDIPVYSEGQVQKQIYVNA 176
/ Db 121 KLA---RHYOMGFETALVCADTFRAGAFDQKQATKAKIPYTGSLTQDPAIVAAG 176
/ QY 177 LKHAKEEHLDFVIDAGRLHIDEALMNELEKVEIAKPEIMLVVDSMTGQDAVNAES 236
/ Db 177 VAKFKKKEPFIIVDSGRHKEEELFTEMTQIQTAVTPDQILVLDSTIGQAAEAQSSA 236
/ QY 237 FDDQLDVTGVTLLKDGDRGGAALSRSVTKPKIKVGVSEKLDGLFLFHPERMASRI 296
/ Db 237 FKATADFGAIIITKTGHAAGGGAISAVAATHPIIYLTGTGHELMDLERFPKAFIOKLL 296
/ QY 297 GNGDVLSLIEKAOQ-DVDOEKAKDLEKKMRESSFTLDDFLEQLDQVKNLGLDDIMKMP 355
/ Db 297 GNGDMAGLVEHVQAVTKDSASAKETKHISEGTYTLDFRENITSIMKAGPLSKLSMIP 356
/ QY 356 GN-NMKGDLKNNSEKQIDHIKAIQSWTPAERNPDITLVN---SRKKRIAGSGRSLSQ 411
/ Db 357 GLSNLTAGLDDDEGSMK-LRRMIYIFDSMTAAELDGDGKMFVEQPSRMVRIACSGTIVR 415
/ QY 412 EVNRLMKQFNDMKMMKQFTGGGKKGKGRNQGNMLKGN 452
/ Db 416 EYEDLLSQHRMAGMAKRV--GGQKKQMR--AQNMLKGN 452
/
/ RESULT 8
/ US-08-317-401E-4
/ Sequence 4, Application US/08317401E
/ Patent No. 5922561
/ GENERAL INFORMATION:
/ APPLICANT: Thompson, Sheryl Ann
/ APPLICANT: Yaver, Debbie Sue
```

TYPE: PRT
OPERATION: Standby] occurs on idermidic

US-09-134-001C-4817

Query Match 18.8%; Score 426.5; DB 4; Length 415;
Best Local Similarity 31.4%; Pred. No. 2.8e-29;
Matches 103; Conservative 76; Mismatches 122; Indels 27; Gaps

QY 8 ERLOATMOKRGKGLTADIKIMREVRLLAFADYNFKVKEFIKTVSERALGSDVMQ 67
DB 100 QNFQEQNNLIARVYKVDDEFFALEEM---LITADVGFNTV---MTLELREAOQR 149
QY 68 SLTPGQOVI-KIV-----QDELTKLGGENTSINMSKPPVVMVGLGAGKTTTAGKL 121
DB 157 TEDLRVIVEKIVEIYHQDDSEAMNIEDRLN-----VILMVGNGVGTITIGKL 209
QY 122 ALLMRKYNKPMVAADYIPPAINQLQTVGKQIDIPYSEGQVQKPOQIVTNALKHAK 181
DB 210 AYRQOE-GRKVMLAGDTFRAGAIQQLNWMGERVGVSONGSDPAAVYDAINAAK 268
QY 182 EHLDFVIIDTAGRLHIDEALMNLKEVKEIAK-----PNEIMLVVDSMTGQDAVNAE 235
DB 269 NKGVDIICDTAGRLQNKSNLMQELDKMKRVINRAIPDAPHEALLCIDATTGNALSQAR 328
QY 236 SFDDQDVTGVLTKLDGTGGAALSIRSVTKPIKFGVSEKLDGLLELPHPERMASRI 295
DB 329 SFKEVINVGIVLTKLDGTAGGIVLAIRNELHIPVKYVGLGKMDLOLPENPE---SYV 385
QY 296 LGM-GDVLSLTEKAQDVDOEK 322
DB 386 YGLFADMEQNEIDPEISRNSSVESEE 413

RESULT 11

US-09-007-476-2
; Sequence 2, Application US/09007476
; Patent No. 6159949

GENERAL INFORMATION:
; APPLICANT: Black, Michael T.

TITLE OF INVENTION: No. 6159949el Ftsy

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/007,476

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GMI0079

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 416 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-007-476-2

Query Match 18.5%; Score 421; DB 4; Length 416;
Best Local Similarity 31.4%; Pred. No. 8.4e-29;
Matches 101; Conservative 78; Mismatches 121; Indels 22; Gaps

QY 8 ERLOATMOKRGKGLTADIKIMREVRLLAFADYNFKVKEFIKTVSERALGSDVMQ 67
DB 97 QNFQEQNNLIARVYKVDDEFFALEEM---LITADVGFNTV---MTLELREAOQR 149
QY 68 SLTPGQOVI-KIV-----QDELTKLGGENTSINMSKPPVVMVGLGAGKTTTAGKL 124
DB 150 NIQDTELRVIVEKIVEIYHQDDSEAMNIEDRLN-----VILMVGNGVGTITIGKL 208
QY 125 MRKYNKPMVAADYIPPAINQLQTVGKQIDIPYSEGQVQKPOQIVTNALKHAKHEE 184
DB 209 RYKMGKVMLAGDTFRAGAIQQLNWMGERVGVDSIQSEGSDPAAVYDAINAAKNG 269
QY 185 LDFVIIDTAGRLHIDEALMNLKEVKEIAK-----PNEIMLVVDSMTGQDAVNAE 238
DB 269 VDILICDTAGRLQNKSNLMQELDKMKRVINRAIPDAPHEALLCIDATTGNALSQAR 328
QY 239 DQDQDVTGVLTKLDGTGGAALSIRSVTKPIKFGVSEKLDGLLELPHPER-----MASR 294
DB 329 EVTNVVGIVLTKLDGTAGGIVLAIRNELHIPVKYVGLGKMDLOLPENPE---SYV 388
QY 295 ILGMGDVLSLTEKAQDVDOEK 316
DB 389 MIEQNEIDITVEN-DQIVTEBK 409

RESULT 12

US-08-986-963-2

; Sequence 2, Application US/08986963

; Patent No. 5958730

GENERAL INFORMATION:

APPLICANT: Rosteck Jr., Paul R.

TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence

TITLE OF INVENTION: Ftsy

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company

STREET: Lilly Corporate Center

CITY: Indianapolis

STATE: Indiana

COUNTRY: US

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/986,963

FILING DATE: December 8, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Webster, Thomas D.

REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: X-11763

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317/276-3334

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 425 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-986-963-2

Query Match 17.5%; Score 397.5; DB 2; Length 425;
Best Local Similarity 32.5%; Pred. No. 9.8e-27;
Matches 105; Conservative 64; Mismatches 109; Indels 45; Gaps

QY 2 AFEGLSERLOATMOKMR-GKGLTEADIKIMREVRALFEADVNFKVY-KEFIKTVSER 59
 Db 116 AETVQEKYDRSLAKTRUGF-----ARUNAFPA--NFRSDVEEFEELELLI 160
 QY 60 ALGSDY-----MQSLTPGQOVKIVODELTCLM---GGENTSINMSKPP 101
 Db 161 LIMSDBGVQVNASNTEELRYEAKLENAKPPDALRRVITIEKLVYEDKGSYDSHFQDN 220
 QY 100 PPTVYMWGLQAGAGTTTAGKIALLMKKYKPKMLVAADYRPAAINOLQTVGKOIDIP 159
 Db 221 -LTVMLFVGNGVGTTSIGKLA-HRYKRTGKKVMLVAADTFRAGAVAQLAEWGRVDVP 278
 QY 160 VTSBGQVQPOQIVYNALKHAKKEHLDVFIIDTAGRLHIDEALMNLKEVKEIAK----- 214
 Db 279 VVTGPKADPASVYFDGNERAVAEGIDILMIDTAGLQNKDNLMAELEKIGRIIKRVVPE 338
 QY 215 -PREIMLVDSMTGODAVNAESFDDQDLDTGVTLTKLDGDTGGAALSIRSVTQKPIKPV 273
 Db 339 APHEFLAIDASTGONALVQAKESKITPLTGIVLTGIDGTARGGVVLAIREELNIPVKL 398
 QY 274 VGMSEKLDGLELPHPERMASRIL 296
 Db 399 IGFEKIDDIGEFENSENFMKGLL 421

RESULT 13

US-09-007-484-2
 ; Sequence 2, Application US/09007484
 ; Patent No. 6072032
 ; GENERAL INFORMATION:
 ; APPLICANT: Black, Michael T.
 ; TITLE OF INVENTION: No. 6072032el Ftsy
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dechert Price & Rhoads
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: PastSEO for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/007,484
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Falk, Stephen T
 ; REGISTRATION NUMBER: 36,795
 ; REFERENCE/DOCKET NUMBER: GMI0081
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-994-2488
 ; TELEFAX: 215-994-2222
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 430 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-09-007-484-2

Query Match 17.5%; Score 397.5; DB 3; Length 430;
 Best Local Similarity 32.4%; Pred. No. 9,9e-27;
 Matches 104; Conservative 64; Mismatches 108; Indels 45; Gaps 9;

QY 4 EGLSERLOATMOKMR-GKGLTEADIKIMREVRALFEADVNFKVY-KEFIKTVSERAL 61

Db 123 ETVQEKYDRSLAKTRUGF-----ARUNAFPA--NFRSDVEEFEELELLI 167
 QY 62 GSDY-----MQSLTPGQOVKIVODELTCLM---GGENTSINMSKPP 101
 Db 168 MSDVGVQVNASNTEELRYEAKLENAKPPDALRRVITIEKLVYEDKGSYDSHFQDN-L 226
 QY 102 PPTVYMWGLQAGAGTTTAGKIALLMKKYKPKMLVAADYRPAAINOLQTVGKOIDIPVY 161
 Db 227 TVMLFVGNGVGTTSIGKLA-HRYKRTGKKVMLVAADTFRAGAVAQLAEWGRVDVPV 285
 QY 162 SEGQVQVQPOQIVYNALKHAKKEHLDVFIIDTAGRLHIDEALMNLKEVKEIAK-----P 215
 Db 286 TQPEKADPASVYFDGNERAVAEGIDILMIDTAGLQNKDNLMAELEKIGRIIKRVVPEAP 345
 QY 216 NEIMLVDSMTGODAVNAESFDDQDLDTGVTLTKLDGDTGGAALSIRSVTQKPIKPVG 275
 Db 346 HETFLAIDASTGONALVQAKESKITPLTGIVLTGIDGTARGGVVLAIREELNIPVKLIG 405
 QY 276 MEEKLDGLELPHPERMASRIL 296
 Db 406 FGEEKIDDIGEFENSENFMKGLL 426

RESULT 14

US-09-309-682-2
 ; Sequence 2, Application US/09309682
 ; Patent No. 6214348
 ; GENERAL INFORMATION:
 ; APPLICANT: Black, Michael T.
 ; TITLE OF INVENTION: No. 6214348el Ftsy
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dechert Price & Rhoads
 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEO for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/309,682
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/007,484
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Falk, Stephen T
 ; REGISTRATION NUMBER: 36,795
 ; REFERENCE/DOCKET NUMBER: GMI0081
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-994-2488
 ; TELEFAX: 215-994-2222
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 430 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-09-309-682-2

Query Match 17.5%; Score 397.5; DB 4; Length 430;
 Best Local Similarity 32.4%; Pred. No. 9,9e-27;
 Matches 104; Conservative 64; Mismatches 108; Indels 45; Gaps 9;

QY 4 EGLSERLOATMOKMR-GKGLTEADIKIMREVRALFEADVNFKVY-KEFIKTVSERAL 61

Db 123 ETVQEKYDRSLKTRTGF-----ARLNAPFA--NFRSVDEEFEELEELLI 167
QY 62 GSDV-----MQSLTPGQQVIVQDELTKLM---GGENTSINMSKPP 101
Db 168 MSDVGQVVASNLTELRYEAKLENKPPDALRRVIEKLVELYEKDGSYDESIHFQDN-L 226
QY 102 TVVMVGLQAGKTTTAGKIALLMRKKNKPKMLVAADIYRPAAINQLQTVGKQIDIPVY 161
Db 227 TVMLFVGNGVGTTSIGKLA-HRYKRTGKKVMLVAADTFRAGAVAQLAEWGRRVDVPV 285
QY 162 SGGDOVKPOQIVTNALKHAKHEHLDFVILDTAGRLHIDEALMELKEVETAK-----P 215
Db 286 TGPERADPASVVDGMRERAVAGSIDILMIDTAGLQNKDMLMAELEKIGRIIRKRVVPEAP 345
QY 216 NEIMLVDSMTGQDANVAESFDDQLDVGTALTKLDGTRGGAALSIRSVTQKPIKTVG 275
Db 346 HETFLALDASGQNALVQAKESKITPLTGIVLTKIDGTARGGVVLAIREELNIPVKLIG 405
QY 276 MSEKLDGLELHHPERMASRIL 296
Db 406 FGEKIDDIGEFNSFNFMKGILL 426

RESULT 15

US-09-007-484-4
; Sequence 4, Application US/09007484
; Patent No. 6072032
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6072032el Ftsy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,484
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-007-484-4

Query Match 17.2%; Score 390; DB 3; Length 277;
Best Local Similarity 33.5%; Pred. No. 2.4e-26;
Matches 92; Conservative 59; Mismatches 96; Indels 28; Gaps 5;
QY 48 VVKEFIKTVGERALGSDV-----MQSLTPGQQVIVQDELTKLM--- 87
Db 1 VDEEFFEELLLNSDVGQVVASNLTELRYEAKLENKPPDALRRVIEKLVELYEKD 60

QY 88 GGENTSINMSKPPTVVMVGLQAGKTTTAGKIALLMRKKNKPKMLVAADIYRPAAIN 147
Db 61 GSYDESIHFQDN-LTVMLFVGNGVGTTSIGKLA-HRYKRTGKKVMLVAADTFRAGAVD 118
QY 148 QLOTVGKQIDIPVYSEGDOVKPOQIVTNALKHAKHEHLDFVILDTAGRLHIDEALMSELK 207
Db 119 QLAEWGRRVDVPVVTGPEKADPASVVDGMRERAVAGSIDILMIDTAGRLQNKDNLMAELE 178
QY 208 EYKEIAK-----PNEIMLVDSMTGQDANVAESFDDQLDVGTALTKLDGTRGGAAL 261
Db 179 KIGRIIRKRVVPEAPHETFLALDASTGQNALVQAKESKITPLTGIVLTKIDGTARGGVVL 238
QY 262 SIRSVTQKPIKTVGMSKLDGLELHPERMASRIL 296
Db 239 AIREELNIPVKLIGFGEKIDDIGEFNSFNFMKGILL 273

Search completed: February 21, 2003, 15:19:52
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 15:19:07 ; Search time 13 seconds
(without alignments)
1087.444 Million cell updates/sec

Title: US-09-943-108a-2

Perfect score: 2274

Sequence: 1 MAFGLSERLQATMKRGK.....GKKGRKQNMKGLNLPF 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31059816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/FCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2261	99.4	455	10	US-09-815-242-13066
2	2235	98.3	455	10	US-09-815-242-5420
3	2235	98.3	455	10	US-09-815-242-12128
4	1418.5	62.4	472	10	US-09-815-242-4945
5	1418.5	62.4	477	10	US-09-815-242-10861
6	1307	57.5	523	10	US-09-815-242-13254
7	1306	57.4	466	10	US-09-815-242-13525
8	1297	57.0	523	10	US-09-814-041A-2
9	1168	51.4	547	9	US-09-738-626-5755
10	1164.5	51.2	453	10	US-09-815-242-10263
11	1150	50.6	457	10	US-09-815-242-11948
12	1138	50.0	462	10	US-09-815-242-10965
13	1127.5	49.6	501	10	US-09-815-242-14089
14	854	37.6	448	10	US-09-815-242-11562
15	432	19.0	497	10	US-09-815-242-10378
16	418	18.4	416	10	US-09-815-242-12126
17	416	18.3	316	10	US-09-815-242-5298
18	404.5	17.8	442	10	US-09-815-242-10904
19	404	17.8	455	10	US-09-815-242-11781

20	403.5	17.7	313	10	US-09-815-242-4992
21	397.5	17.5	425	10	US-09-815-242-13263
22	397.5	17.5	430	10	US-09-827-663-2
23	396.5	17.4	429	10	US-09-815-242-13589
24	395.5	17.4	414	10	US-09-815-242-11094
25	390	17.2	277	10	US-09-827-663-4
26	376	16.5	510	9	US-09-738-626-5759
27	151	6.6	1945	9	US-09-927-597-2
28	151	6.6	1979	9	US-09-927-597-4
29	140.5	6.2	2025	10	US-09-815-242-5703
30	140.5	6.2	3158	10	US-09-815-242-12611
31	134	5.9	705	10	US-09-815-242-5334
32	134	5.9	713	10	US-09-815-242-12322
33	132	5.8	856	10	US-09-815-242-11310
34	130.5	5.7	1786	9	US-09-742-096-3
35	129.5	5.7	2437	10	US-09-815-242-5834
36	129.5	5.7	6281	10	US-09-815-242-12996
37	129	5.7	856	10	US-09-815-242-11489
38	127.5	5.6	866	12	US-10-007-693-113
39	126	5.5	1342	10	US-09-815-242-10438
40	124.5	5.5	800	10	US-09-815-242-5349
41	124.5	5.5	800	10	US-09-815-242-12139
42	124.5	5.5	800	10	US-09-815-242-13136
43	124.5	5.5	1212	10	US-09-815-242-5279
44	124.5	5.5	1482	10	US-09-815-242-13484
45	121.5	5.3	592	10	US-09-861-451A-72

ALIGNMENTS

RESULT 1

- US-09-815-242-13066
- Sequence 13066, Application US/09815242
- Patent No. US20020061569A1
- GENERAL INFORMATION:
- APPLICANT: Haselbeck, Robert
- APPLICANT: Ohlsen, Karl L.
- APPLICANT: Zyskind, Judith W.
- APPLICANT: Wall, Daniel
- APPLICANT: Trawick, John D.
- APPLICANT: Carr, Grant J.
- APPLICANT: Yamamoto, Robert T.
- APPLICANT: Xu, H. Howard

- TITLE OF INVENTION: Identification of Essential Genes in
- TITLE OF INVENTION: Prokaryotes
- FILE REFERENCE: ELITRA.01A
- CURRENT APPLICATION NUMBER: US/09/815,242
- CURRENT FILING DATE: 2001-03-21
- PRIOR APPLICATION NUMBER: 60/191,078
- PRIOR FILING DATE: 2000-03-21
- PRIOR APPLICATION NUMBER: 60/206,848
- PRIOR FILING DATE: 2000-05-23
- PRIOR APPLICATION NUMBER: 60/207,727
- PRIOR FILING DATE: 2000-05-26
- PRIOR APPLICATION NUMBER: 60/242,578
- PRIOR FILING DATE: 2000-10-23
- PRIOR APPLICATION NUMBER: 60/253,625
- PRIOR FILING DATE: 2000-11-27
- PRIOR APPLICATION NUMBER: 60/257,931
- PRIOR FILING DATE: 2000-12-22
- PRIOR APPLICATION NUMBER: 60/269,308
- PRIOR FILING DATE: 2001-02-16
- NUMBER OF SEQ ID NOS: 14110
- SOFTWARE: FastSeq for Windows Version 4.0
- SEQ ID NO 13066
- LENGTH: 455
- TYPE: PRT
- ORGANISM: Staphylococcus aureus

US-09-815-242-13066

Query Match 99.4%; Score 2261; DB 10; Length 455;
Best Local Similarity 99.3%; Pred. No. 1.3e-154;

Matches	452; Conservative	2; Mismatches	1; Indels	0; Gaps	0;
QY	1	MAFEGLSRLQATMOKMRGKGLTADIKIMREVRILALFEADVNFVKVYKFSFIKYSERA	60		
Db	1	MAFEGLSRLQATMOKMRGKGLTADIKIMREVRILALLEADVNFVKVYKFSFIKYSERA	60		
QY	61	LGSVDYMQSLTPGQOVIKIVQDELTKLNGENTSNINSNKPPVVMVGLQAGAKTTTAGK	120		
Db	61	LGSVDYMQSLTPGQOVIKIVQDELTKLNGENTSNINSNKPPVVMVGLQAGAKTTTAGK	120		
QY	121	LALLMRKKYNNKPMVAAIDYTPAAINQLQVGRQIDIPVYSEGQVQPOQIVTNALKHA	180		
Db	121	LALLMRKKYNNKPMVAAIDYTPAAINQLQVGRQIDIPVYSEGQVQPOQIVTNALKHA	180		
QY	181	KEEHLDFVIDTAGRLHIDEALMNLKEVKVKIAKPEIMLVVDSMTGQADVNVAESFDQ	240		
Db	181	KEEHLDFVIDTAGRLHIDEALMNLKEVKVKIAKPEIMLVVDSMTGQADVNVAESFDQ	240		
QY	241	LDVTVITLKLGDTRGGAALSIKRVSTQKPIKFGVSEKLDGLLELHFPERWASRIKMGD	300		
Db	241	LDVTVITLKLGDTRGGAALSIKRVSTQKPIKFGVSEKLDGLLELHFPERWASRIKMGD	300		
QY	301	VLSLEKAQDVQDEKAKOLEKKMRESEFTLDDFLQLDQVKNLGPLDDIMKMTPGMNKX	360		
Db	301	VLSLEKAQDVQDEKAKOLEKKMRESEFTLDDFLQLDQVKNLGPLDDIMKMTPGMNKX	360		
QY	361	KGLDKLNNSERQIDHILAKIIQSMTPAERNPDTILNYSRKKRTAGSGSLOEVNRLMKOF	420		
Db	361	KGLDKLNNSERQIDHILAKIIQSMTPAERNPDTILNYSRKKRTAGSGSLOEVNRLMKOF	420		
QY	421	NDMKKMMKQFTGGGKGKKGRKNOMNMLKGNLPF	455		
Db	421	NDMKKMMKQFTGGGKGKKGRKNOMNMLKGNLPF	455		

```

RESULT 2
US-09-815-242-5420
; Sequence 5420, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5420
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

```

Query Match	98.3%;	Score 2235;	DB 10;	Length 455;
Best Local Similarity	98.28;	Pred. No. 9,3e-153;		
Matches 447;	Conservative 3;	Mismatches 15;	Indels 0;	Gaps
QY 1	MAFEGLSERLQATMQKMRGKGLFEADIKIMRERYLALFEADVNFVKYKEFIKTVSERA	60		
Db 1	MAFEGLSERLQATMQKMRGKGLFEADIKIMRERYLALFEADVNFVKYKEFIKTVSERA	60		
QY 61	LGSDWQSLPFGQGVKIVQDELITKMGENTSIINMSNKPPTVVMVWGLQGAGKTTAGK	120		
Db 61	LGSDWQSLPFGQGVKIVQDELITKMGENTSIINMSNKPPTVVMVWGLQGAGKTTAGK	120		
QY 121	LALLMRKKYNNKPEMLVAADIYRPAAINQLOTVGKQIDIPVYSGSDQVKPQOLVTVNALKHA	180		
Db 121	LALLMRKKYNNKPEMLVAADIYRPAAINQLOTVGKQIDIPVYSGSDQVKPQOLVTVNALKHA	180		
QY 181	KEEHLDFVIIDTAGRLHIDEALNNELKEVKEIAKPEINMLVDSMTGQDANVAESFDDQ	240		
Db 181	KEEHLDFVIIDTAGRLHIDEALNNELKEVKEIAKPEINMLVDSMTGQDANVAESFDDQ	240		
QY 241	LDVTGVTLTKLGDTRGGAALSTRSVTKPIKFGVGNSEKLDGLELPHPERMASRILGMD	300		
Db 241	LDVTGVTLTKLGDTRGGAALSTRSVTKPIKFGVGNSEKLDGLELPHPERMASRILGMD	300		
QY 301	VLSLIERAQODVDQEKADLEKKMRSSFTLDDFLEQLDQVKNLGPLDDIMKMTPGNKM	360		
Db 301	VLSLIERAQODVDQEKADLEKKMRSSFTLDDFLEQLDQVKNLGPLDDIMKMTPGNKM	360		
QY 361	KGLDKLNMSKQIDHIKAIQSMTPAERNNPDTLVNSKKRIAKGSGRSLOEVNRLMKQF	420		
Db 361	KGLDKLNMSKQIDHIKAIQSMTPAERNNPDTLVNSKKRIAKGSGRSLOEVNRLMKQF	420		
QY 421	NDMKKMMKQFTGGGKKGKGNQNMNLKGMNLFP	455		
Db 421	NDMKKMMKQFTGGGKKGKGNQNMNLKGMNLFP	455		

```

RESULT 3
US-09-815-242-12128
; Sequence 12128, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Onlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0

```

; SEQ ID NO 12128

; LENGTH: 455

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12128

Query Match

Best Local Similarity 98.3%; Score 2235; DB 10; Length 455;

Matches 447; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAPEGLSERLOATQWQMRGKGLTEADIKIMREVRALALFEADVNFVKVKEPIKTVSERA 60

Db 1 MAPEGLSERLOATQWQMRGKGLTEADIKIMREVRALALFEADVNFVKVKEPIKTVSERA 60

QY 61 LGSVDVMSLTQGGQVIRIVODELTCLMGGENTSINMSNKPPTVVMVGLQGAGKTTTAGK 120

Db 61 LGSVDVMSLTQGGQVIRIVODELTCLMGGENTSINMSNKPPTVVMVGLQGAGKTTTAGK 120

QY 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPVYSGDQVQKQOIVTNALKHA 180

Db 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPVYSGDQVQKQOIVTNALKHA 180

QY 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPNELVDSMTGQDAVNVAESFDDQ 240

Db 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPNELVDSMTGQDAVNVAESFDDQ 240

QY 241 LDVGTGVTLLKLDGDTGGAALSIRSVTQKPIKFGVMSKLDGLLELPHPERMASRIILMGD 300

Db 241 LDVGTGVTLLKLDGDTGGAALSIRSVTQKPIKFGVMSKLDGLLELPHPERMASRIILMGD 300

QY 301 VLSLEKAQDVDEKAKOLEKKEKRESFTLDDLEQLDQVKNLGPDDIMKMPGMNMK 360

Db 301 VLSLEKAQDVDEKAKOLEKKEKRESFTLDDLEQLDQVKNLGPDDIMKMPGMNMK 360

QY 361 KGLDKLNMSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAGSGRSLSQEVNRLMKQF 420

Db 361 KGLDKLNMSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAGSGRSLSQEVNRLMKQF 420

QY 421 NDMMKMKMKTFTGGGKGGKGRNOMNMLKGNLFP 455

Db 421 NDMMKMKMKTFTGGGKGGKGRNOMNMLKGNLFP 455

RESULT 4

US-09-815-242-4945

; Sequence 4945; Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4945

; LENGTH: 472

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-815-242-4945

Query Match

Best Local Similarity 62.4%; Score 1418.5; DB 10; Length 472;

Matches 273; Conservative 86; Mismatches 89; Indels 15; Gaps 2;

QY 1 MAPEGLSERLOATQWQMRGKGLTEADIKIMREVRALALFEADVNFVKVKEPIKTVSERA 60

Db 1 MAPEGLSERLOATQWQMRGKGLTEADIKIMREVRALALFEADVNFVKVKEPIKTVSERA 60

QY 61 LGSVDVMSLTQGGQVIRIVODELTCLMGGENTSINMSNKPPTVVMVGLQGAGKTTTAGK 120

Db 61 LGSVDVMSLTQGGQVIRIVODELTCLMGGENTSINMSNKPPTVVMVGLQGAGKTTTAGK 120

QY 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPVYSGDQVQKQOIVTNALKHA 180

Db 121 LALLMRKYNKPMVAADYIRPAAINOLQTVGKQIDIPVYSGDQVQKQOIVTNALKHA 180

QY 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPNELVDSMTGQDAVNVAESFDDQ 240

Db 181 KEHLDVFIIDTAGRLHIDEALMNEKEVKEIAKPNELVDSMTGQDAVNVAESFDDQ 240

QY 241 LDVGTGVTLLKLDGDTGGAALSIRSVTQKPIKFGVMSKLDGLLELPHPERMASRIILMGD 300

Db 241 LDVGTGVTLLKLDGDTGGAALSIRSVTQKPIKFGVMSKLDGLLELPHPERMASRIILMGD 300

QY 301 VLSLEKAQDVDEKAKOLEKKEKRESFTLDDLEQLDQVKNLGPDDIMKMPGMNMK 360

Db 301 VLSLEKAQDVDEKAKOLEKKEKRESFTLDDLEQLDQVKNLGPDDIMKMPGMNMK 360

QY 361 KGLDKLNMSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAGSGRSLSQEVNRLMKQF 420

Db 361 KGLDKLNMSEKQIDHIKAIQSMTPAERNPDTLNVSRKKRIAGSGRSLSQEVNRLMKQF 420

QY 421 NDMMKMKMKTFTGGGKGGKGRNOMNMLK 449

Db 421 NDMMKMKMKTFTGGGKGGKGRNOMNMLK 449

RESULT 5

US-09-815-242-10861

; Sequence 10861; Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 10861
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10861
```

```
Query Match          62.4%; Score 1418.5; DB 10; Length 477;
Best Local Similarity 59.08; Pred. No. 3e-94; Mismatches 89; Indels 15; Gaps 2;
Matches 273; Conservative
```

```
QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRALFADVNFKVKEFKTYSERA 60
|||||: ||| | :|: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 1 MAFESLTNRLLQAMGKIRKKGKSEADVKEKMRREIRLALLEADVNLQVYKDFTKVRERA 60

QY 61 LGSVMQSLTPGQOVIKIVQDELTKMGCGTISNNMKNPPTVVMVVGLOGAGKTTTACK 120
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 61 VGVFVLESUSPAQIIVKIVDELTGKISVETVNLKSPKPTVIMTGLGAGKTTTACK 120

QY 121 LALIMRKNKPKMELVAADIYRPAAINQLQTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 121 LAKHLKMTNARPLLIAGDVYRPAIDQLKVLQQLLEVYDFMGTDANPVEIVRQGLALA 180

QY 191 KEELDFVIDTAGRLHIDEALMNELKEVKEIAKNEIMLVVDSMTGQADVNAESFDDQ 240
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 181 KEKKNYDVIDTAGRLHIDEALMDELQIKELANPNEILLVVDAMTGDADVNAESFDDQ 240

QY 241 LDVTGVTLLKLDGDRGGAALSIRSVTKPIKTVGSEKLDGLLELHFPERMASRIILMGD 300
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 241 LGITGVYTKLDGDRGGAALSIRAVTGAPIKFVGSEKLTDLLELHFPDRMSRIILMGD 300

QY 301 VLSLIEKAQDVQDKAKDLEKKMRSEFTLDDFLEQLDVKNLGPDLDDIMKMPGNMK 360
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 301 MLTIEKAQDVQDKAKDELAQKMKENSEDFNDFIEQLDVQVGMGPIEDLLKMPGNM 360

QY 361 KGLDKLMSKQZIDHKAIIOSMTPAERNPDPLNVSRKKRIAGSGESLOEVNRLMKQF 420
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 361 PGIENTVQPKDVARKRAVLSMTPAERNPDPLNPSRRRIAGSGSVNVVNRMIKQF 420

QY 421 NDMKKMKMOKFT-----GGGKGGKGGKGNQNMNLMK 449
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 421 KESKMMQMSKGMNIFGMQMLGGGVKGLGKM-AMNRMK 462
```

```
RESULT 6
US-09-815-242-13254
; Sequence 13254, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13254
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13254

Query Match          57.5%; Score 1307; DB 10; Length 523;
Best Local Similarity 55.1%; Pred. No. 3.3e-86;
Matches 253; Conservative 90; Mismatches 102; Indels 14; Gaps 2;

QY 1 MAFEGLSERLQATMOKMRGKGLTEADIKIMREVRALFADVNFKVKEFKTYSERA 60
|||||: ||| | :|: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 1 MAFESLTNRLLQAMGKIRKKGKISDVQEQATKEIRLALLEADVNLQVYKDFTKVRERA 60

QY 61 LGSVMQSLTPGQOVIKIVQDELTKMGCGTISNNMKNPPTVVMVVGLOGAGKTTTACK 120
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 61 VGVFVLESUSPAQIIVKIVDELTGKISVETVNLKSPKPTVIMTGLGAGKTTTACK 120

QY 121 LALIMRKNKPKMELVAADIYRPAAINQLQTVGKQIDIPVYSEGQVQKPOQIVTNALKHA 180
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 121 LANKLKEENARPLVAADIYRPAIDQLKVLQQLLEVYDFMGTDANPVEIVRQGLEQA 180

QY 181 KEELDFVIDTAGRLHIDEALMNELKEVKEIAKNEIMLVVDSMTGQADVNAESFDDQ 240
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 181 QTNEDYVIDTAGRLHIDEALMDELQIKELANPNEILLVVDAMTGDADVNAESFDDQ 240

QY 241 LDVTGVTLLKLDGDRGGAALSIRSVTKPIKTVGSEKLDGLLELHFPERMASRIILMGD 300
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 241 LEVTVVLLKLDGDRGGAALSIRVHTGPKETGTGKIDTIEFHPDRMSRIILMGD 300

QY 301 VLSLIEKAQDVQDKAKDLEKKMRSEFTLDDFLEQLDVKNLGPDLDDIMKMPGNMK 360
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 301 MLTIEKAQDVQDKALEMAEKRENTDFNDFIDQLDVQVGMGPIEDLLKMPGNM 360

QY 361 KGLDKLMSKQZIDHKAIIOSMTPAERNPDPLNVSRKKRIAGSGESLOEVNRLMKQF 420
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 361 PAFQNMKVDQRIARKRAIVSSNTPPEERNPDPLNPSRRRIAGSGNTFVEVNRKIDF 420

QY 421 NDMKKMKMOKFTGGGKGGKGGKGNQNMNLMKGN-----NLP 454
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
Db 421 NOAKQLMGMVSS-----DNKMKMKQMGINPNLNP 450
```

```
RESULT 7
US-09-815-242-13525
; Sequence 13525, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
```



```

; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent In ver. 3.0
; SEQ ID NO 5755
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5755

Query Match          51.4%; Score 1168; DB 9; Length 547;
Best Local Similarity 49.1%; Pred. No. 3.2e-76;
Matches 237; Conservative 85; Mismatches 125; Indels 36; Gaps 5;

Qy 3 FEGLSERLOATMOKRGKGLTEADIKIMREVRRLALFEADVNFVKVKEFIKTVSERALG 62
Db 2 FESLSRLNSALSGLRGKGLTEADINATREIRLALLEADVSLTVVRAFINRIKERAVG 61

Qy 63 SDYMSQSLTPGQOVIQVDELTKIMGGENTSINMSKPPTVMMVGLQGAGKTTAGKLA 122
Db 62 AEYSQALNPAQOQVIKIVNEELVQLGGETRRSLAKNPPTVINLAGLQGAGKTTAGKLS 121

Qy 123 LLMRRKYNKKPMLVAADIYRPAALNOLQTVGKQIDIPVY-----SEGD 165
Db 122 KHLVKQ-GHTPMLVACDLQRGAVQQLQIVGERAGVTTFADPDGTSIDSLEHEKGTSHGD 180

Qy 166 QVPQOIVYNALKHAKHEHLDVFIIDTAGRLHIDBALMNEIKVEIKAPNEIMLVDSM 225
Db 181 ---PVEVARAGTEEAKRQTHDIVDVTAGRLGIDETLMTQARNIREAINPDEVLFVIDSM 237

Qy 226 TGGDAYNVAESDDLDVTVGTLTKLDGTRGGAALSIRSVTQPIKVEVMSSEKLDGLEL 285
Db 238 IGGDADTAAEPGROGVDFTGVVLTFLKLDGDARGAALSIREVTGKPIINFAGSTGEKLDPEV 297

Qy 286 FHPERMASRLGMDVLSLEKAQDVDEKAKDLEKKMRESSTLDDFLQDLDVKNLG 345
Db 298 FHPERMASRLGMDVLSLEKAQDVDEKAEVAQKLGSGEFLTEDFDQDMLIRRG 357

Qy 346 PLDDIMKMTPGNKKGLDKLNKSEKQIDHIKAIQSMTPAERNPPTLVNRSKKRIAG 405
Db 358 PIGNILKMLPGCKQMSQADN-VDEKQDLRIQAILTGMTPAERNPKNILNASRRKRIAG 416

Qy 406 SGRSLOEVNRLMKQFNNDKMKKQFTG-----GGKKGKGNQNMNKLGM 451
Db 417 SGVTSEVANKUVERFEARKVMQMAQGFQNGFGSRSAATKKQAKGKNGKPKAKGP 476

Qy 452 NLP 454
Db 477 TQP 479

RESULT 10
US-09-815-242-10263
; Sequence 10263, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10263
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10263

Query Match          51.2%; Score 1164.5; DB 10; Length 453;
Best Local Similarity 51.0%; Pred. No. 4.4e-76;
Matches 231; Conservative 94; Mismatches 121; Indels 7; Gaps 2;

Qy 3 FEGLSERLOATMOKRGKGLTEADIKIMREVRRLALFEADVNFVKVKEFIKTVSERALG 62
Db 2 FDLNLTDLRSRTURNISGRGLTDEDNVKTDLREVRMLLEADVLPVVVREFINRVKEKAVG 61

Qy 63 SDYMSQSLTPGQOVIQVDELTKIMGGENTSINMSKPPTVMMVGLQGAGKTTAGKLA 122
Db 62 HEYNKSLTPGQEFYKIVNELVAAMGEENQTLNLAQPPAVVLNAGLQGAGKTTSGKLG 121

Qy 123 LLMRRKYNKKPMLVAADIYRPAALNOLQTVGKQIDIPVYSEGQVKKPQQVTNALHAKHE 182
Db 122 KFLREXKHKKVLVWSADYVPAALQLETLAEQVGVDFPDSVGVQKPDVTYNAALKEAKL 181

Qy 183 EHLDFVILDTAGRLHIDBALMNEIKVEIKAPNEIMLVDSMTGODAVNVAESDDQLD 242
Db 182 KFYDVVLVDVDTAGRLHIDBALMNEIKVEIKAPNEIMLVDSMTGODAVNVAESDDQLD 241

Qy 243 VTGVTTLTKLDGTRGGAALSIRSVTQPIKVEVMSSEKLDGLELPHPERMASRLGMDV 302
Db 242 LTGVVLTKVDGARGAALSIRHITGKPIKFLGVGEKTEALEPFPDRIASRLGMDVL 301

Qy 303 SLEKAQDVDEKAKDLEKKMRE-SSTFLDLDLQDQVKNLGLDQDLMKMTIPGNKMK 361
Db 302 SLEEDIESVRAQAEKSLAKKKGDFDLNDFLQLRQMKNMGMASLMLKGLPGMGQIP 361

Qy 362 GLDLNASEKQIDHIKAIQSMTPAERNPPTLVNRSKKRIAGSGRSLOEVNRLMKQFN 421
Db 362 DNVKSDQDDKVLVRMEALINSMTWKEAKPEIKGSRRRRIAGCGMQGVQDVNRLKQFD 421

Qy 422 DMKKMKKQFTGGGKKGKGNQNMNKLGMNLP 454
Db 422 DMQRMMK-----KMKKGMAKMRSMKGMMP 448

RESULT 11
US-09-815-242-11948
; Sequence 11948, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11948
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11948

```

```

Query Match          50.6%; Score 1150; DB 10; Length 457;
Best Local Similarity 50.5%; Pred. No. 4.9e-75;
Matches 231; Conservative 93; Mismatches 121; Indels 12; Gaps 3;

```

```

QY 3 FEGLSERLOATMQRGKGLTEADIKIMREVRLALFEADVNFVKVKEFTKTVSRALG 62
DB 2 FENLDRSOTLRHTVTKAKLTEDNTKDTLREVRMALLEADVLPVVKDFNKVRAGV 61
QY 63 SDVMQSLTPGQOVTKIVQDELTKLMGGENTSINMSKPPVTVMVVGLOGAGKTTAGKLA 122
DB 62 TEVSLSLTPGQAFKIVRAEELMGAANEDLALSAPPVAVILMAGLOGAGKTTAGKLA 121
QY 123 LMRKKNKPMVAADIVRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 182
DB 122 RFLKERRKKSVMYVADIVRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 181
QY 183 EHLDFVITDTAGRLHIDEALMNELEKVEIKAEKPNELMLVDSMTGQDANVAESFDOLD 242
DB 182 KFLDERKKKLVVVSADIVRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 241
QY 243 VTGVTITKLDGTRGGAALSIRSVTQPIKFGVMSKIDGLLELPHPERMASRILGMDVL 302
DB 242 LTGVVLTQVGDARGGAALSIRSVTQPIKFGVMSKIDGLLELPHPERMASRILGMDVL 301
QY 303 SLIEKAQDVQDQKADLEKKMRE--SSFTLDDFLEQDQVKNLGLPDDIMKPIGMMKAK 358
DB 302 SLIEKAQDVQDQKADLEKKMRE--SSFTLDDFLEQDQVKNLGLPDDIMKPIGMMKAK 361
QY 359 KMKGLDKLNMSEKQIDHKAIIQOSTPAERNPDTLNVSKRKRIAGSGRSLSQEVNRLMK 418
DB 362 LAQMGAQGAEEKQFKQWEALINSMTPEGRDPEMISGRKRRIALGSGTQVQDVGRLIK 421
QY 419 QFNDMKKMKKQFTGGGKGGKGGKKNQNMKMLKMLPP 455
DB 422 QHKQMKMKKKVYTAG-----GMAKMMRGMSNF 450

```

```

RESULT 12
US-09-815-242-10965
; Sequence 10965, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes

```

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10965
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10965

```

```

Query Match          50.0%; Score 1138; DB 10; Length 462;
Best Local Similarity 50.0%; Pred. No. 3.6e-74;
Matches 225; Conservative 93; Mismatches 122; Indels 10; Gaps 2;

```

```

QY 3 FEGLSERLOATMQRGKGLTEADIKIMREVRLALFEADVNFVKVKEFTKTVSRALG 62
DB 2 FENLDRSOTLRHTVTKAKLTEDNTKDTLREVRMALLEADVLPVVKDFNKVRAGV 61
QY 63 SDVMQSLTPGQOVTKIVQDELTKLMGGENTSINMSKPPVTVMVVGLOGAGKTTAGKLA 122
DB 62 EVNKSLSLTPGQEPKIVQRELEKAMGEANESLNLAQPPAVILMAGLOGAGKTTAGKLA 121
QY 123 LMRKKNKPMVAADIVRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 182
DB 122 KFLDERKKKLVVVSADIVRPAALNQLQTVGKQIDIPVYSEGQVKKPQOIVTNALKHAK 181
QY 183 EHLDFVITDTAGRLHIDEALMNELEKVEIKAEKPNELMLVDSMTGQDANVAESFDOLD 242
DB 182 KFYDVLVITDTAGRLHIDEALMNELEKVEIKAEKPNELMLVDSMTGQDANVAESFDOLD 241
QY 243 VTGVTITKLDGTRGGAALSIRSVTQPIKFGVMSKIDGLLELPHPERMASRILGMDVL 302
DB 242 LTGVVLTQVGDARGGAALSIRSVTQPIKFGVMSKIDGLLELPHPERMASRILGMDVL 301
QY 303 SLIEKAQDVQDQKADLEKKMRE--SSFTLDDFLEQDQVKNLGLPDDIMKPIGMMKAK 361
DB 302 SLIEKAQDVQDQKADLEKKMRE--SSFTLDDFLEQDQVKNLGLPDDIMKPIGMMKAK 361
QY 362 GLDKLNMSEKQIDHKAIIQOSTPAERNPDTLNVSKRKRIAGSGRSLSQEVNRLMKQFN 421
DB 362 EHVKNQVDDKFKYKMEALINSMTLKERANFDITKGSRRRIALGSGTQVQDVNKLKQFD 421
QY 422 DMKKMKKQFTGGGKGGKGGKKNQNMKMLKGM 451
DB 422 EMQRMKMKKGG-----MAKMMRGM 442

```

```

RESULT 13
US-09-815-242-14089
; Sequence 14089, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```



```
RESULT 15
US-09-815-242-10378
; Sequence 10378, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITFA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-08-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10378
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10378

Query Match          19.0%; Score 432; DB 10; Length 497;
Best Local Similarity 31.3%; Pred. No. 1.5e-23;
Matches 98; Conservative 72; Mismatches 119; Indels 24; Gaps 5;

QY  4  EGLSERLOATMQRG-----KGLTEADIKIMREYRLALFADVNFKVKEFI 53
Db  193  EGFFARLRSLTKENLGSFISLFRGKKIDDD---LFEELFEQLLIADVGVETTRKII 249

QY  54  KTVSRALGSDVMOSLPPGOVTKIVODELTKMGENTSIKSNKPPTVYVMYVLOGAG 113
Db  250  TWITEGA---SRQLDRAELXGLKBEEMGEILAKVDEPLNVEGKAPFVILMVGVNGVG 305

QY  114  KTTTAGKIALMRKYNKPMVAADIVRPAAINQLQTVGKIDIPVWSEGDQVKPQIV 173
Db  306  KTTTGKILARQFEOQ-CKSVMLAAGTFRAAAVEQLQWVGQORNIPVIAQHTGADSASVI 364

QY  174  TNALKHAKDEHLDPVLDTAGRLHIDEALMMLKEVKEIAK-----PNEIMLVVDSMTG 227
Db  365  FTAIOAKARNIDVLIADTAGRLQNKSHMLELAKIVRYVMKLDVDEAPHEVNLTIIDASTG 424

QY  228  QAVNVAESFDDQDLDTGVTLTKDGTDRGGAALSIRSVTKPIKEVGMSEKLDGLELFH 287
Db  425  QNAVSAQLPHEANGVLGITLTLDGTAKGVVSVADQFGIPYIRVIGYGERIEDLRPFK 484

QY  288  PERMASRILGMCD 300
Db  485  ADDFTALFARED 497
```

Search completed: February 21, 2003, 15:26:44
Job time : 15 secs